University of Glasgow

The role of the transcriptional regulator, DsdC, in *Escherichia coli* pathogenicity

A thesis submitted to the University of Glasgow for the degree of Doctor of Philosophy

Natasha C A Turner BSc (Hons)

Submitted August 2021

Institute of Infection, Immunity, and Inflammation

College of Medical, Veterinary, and Life Sciences

University of Glasgow

Acknowledgements

Firstly, I would like to thank Professor Andrew Roe for all of his support throughout this PhD. I appreciate all of the advice (and cups of tea) he has given, and for introducing me into the fascinating world of bacterial gene regulation. Thank you also to my second supervisor, Professor Konstantinos Gerasimidis, I am grateful for the time I got to spend in your lab.

I have been lucky enough to have a wonderful, supportive group during my PhD, and I would like to thank all members of the Roe group who I have been fortunate to meet and to have shared many brews, laughs, tears, and cake with. A special thank you goes to Dr James Connolly, Dr Nicky O'Boyle, Dr Liyana Azmi, and Jennifer Hallam, whose guidance and support over the last few years have been invaluable.

Doing a PhD has been one of the hardest, but greatest, times of my life. I want to extend my love and gratitude to everybody in the Bacteriology department that has helped me along the way. A special mention must go to Dr Andreas Haag, Aisling Brady, Dr Michael Ormsby, and Dr Khedidja Mosbahi for their help, patience, and assistance in the worlds of bacteriophages, LPS and protein purification. I'd also like to thank Dr Douglas Browning, from the IMI at the University of Birmingham, for charitably taking me under his wing to learn about DNase I footprinting and the world of radiation. Many thanks must also go to the University of Glasgow MVLS DTP programme for funding my studentship.

To my parents; Carol and Dave. You have always been my biggest supporters, sacrificing so much to get me to where I am today. As your annoyingly daft, laal bébé, I will always be thankful and appreciate the efforts you have made. Also, to my Nanna Chris, thank you for always inspiring me to achieve my greatest. Finally, to Alex. Thank you for your unwavering support during the last several years, but especially during the last few months. Without you, this time wouldn't have been half as fun.

As we say back yam; here's to the future, marras.

Table of contents

List of Figures

List of Tables

Abbreviations

α-CTD - α-C-terminal domain

α-NTD - α-N-terminal domain

ABS – activation binding site

AMR – antimicrobial resistance

APEC – avian pathogenic *Escherichia coli*

BCA - bicinchoninic acid

BSA – bovine serum albumin

ChIP – chromatin immunoprecipitation

ChIP-Seq – chromatin immunoprecipitation coupled with next generation sequencing

CIP – calf intestinal phosphatase

CFU – colony forming units

CNF1 - cytotoxic necrotising factor 1

CNS – central nervous system

CRC – colorectal cancer

CRP – catabolite repressor protein

CSF – cerebral spinal fluid

CV – colume volume

DAO – D-amino acid oxidase

DAOA – D-amino acid oxidase activator

DAEC – diffusely adherent *Escherichia coli*

ddH_2O - distilled deionised water

DEG – differentially expressed gene

DGE – differential gene expression

DMEM - Dulbecco's Minimal Eagle Medium

DSB – double strand breaks

dsDNA – double stranded DNA

EAEC – enteroaggregative *Escherichia coli*

EDTA - ethylenediaminetetraacetic acid

EHEC – enterohaemorrhagic *Escherichia coli*

EIEC – enteroinvasive *Escherichia coli*

ENA – European nucleotide archive

EMSA – electrophoretic mobility shift assay

EPEC – enteropathogenic *Escherichia coli*

ETEC – enterotoxigenic *Escherichia coli*

ETT2 – *Escherichia coli* type III secretion system 2

ExPEC – extraintestinal pathogenic *Escherichia coli*

FBS – foetal bovine serum

GO – gene ontology

GST - glutathione S-transferase

HBMEC – human brain microvascular endothelial cells

HTH – helix-turn-helix

HGT – horizontal gene transfer

IBC – intracellular bacterial communities

Ibe – invasion of brain endothelial cells

InPEC – intestinal pathogenic *Escherichia coli*

IPTG - isopropyl ß-D-1-thiogalactopyranoside

LB - Luria-Bertani broth

LEE – locus of enterocyte effacement

LPS – lipopolysaccharide

LTTR – LysR-type transcriptional regulator

MBP – maltose binding protein

MEM-HEPES – Minimal Essential Medium, HEPES modification

MGE – mobile genetic elements

MRSA – methicillin resistant *Staphylococcus aureus*

N-CAM – neural cell adhesion molecule

nfH_2O – nuclease free water

NMDA – *N*-methyl-D-aspartate

NMEC – neonatal meningitis associated *Escherichia coli*

MGE – mobile genetic islands

OMP – outer membrane protein

OmpA – outer membrane protein A

PAI – pathogenicity-associated island

PBS – phosphate buffered saline

PBST - phosphate buffered saline Tween

PCR – polymerase chain reaction

PFA – paraformaldehyde

PFU – plaque forming units

PI – post-infection

PMN - polymorphonuclear neutrophils

PTA – phage top agar

QC – quality control

RBS – regulatory binding site

RNA-Seq – RNA sequencing

RNAP – RNA polymerase

RPc – RNA polymerase closed complex

RPo – RNA polymerase open complex

RT-qPCR – quantitative real time PCR

RTX – repeats-in-toxin

SDS – sodium dodecyl sulphate

SDS-PAGE - sodium dodecyl sulphate-polyacrylamide gel electrophoresis

SEM – standard error of the mean

SEPEC – sepsis-associated *Escherichia coli*

SOC – super optimal broth with catabolite repression

SSM – slipped strand mispairing

ST – sequence type

TAE - Tris-acetate-ethylenediaminetetraacetic acid

TE – Tris- ethylenediaminetetraacetic acid

TBS – Tris buffered saline

TCS – two component systems

TF – transcription factor

TFBS – transcription factor binding site

TRAP – *trp* RNA-binding attenuation protein

TSS – transcriptional start site

T1F – type 1 fimbriae

T2SS – type two secretion system

T3SS – type three secretion system

T6SS – type six secretion system

UPEC – uropathogenic *Escherichia coli*

UTI – urinary tract infection

WT – wild-type

Abstract

Within the mammalian host, bacteria are exposed to a large number of nutritional and chemical stimuli. Modification of gene transcription, in response to environmental signals, enables bacteria to compete and adapt within their preferred niches. D-serine is a host metabolite that has become a focus in recent years due to its diverse roles in neurotransmission and signalling in humans, as well as its unique role in affecting gene expression in bacteria. Indeed, previous work from the Roe laboratory revealed that D-serine played a critical role in controlling the expression of a main virulence determinant in enterohaemorrhagic *Escherichia coli* (EHEC), the type three secretion system (T3SS). Conversely, some commensal and extraintestinal pathogenic *E. coli* (ExPEC) can catabolise D-serine, using it as a carbon source.

Uropathogenic *E. coli* (UPEC) and neonatal meningitis-associated *E. coli* (NMEC) represent two clinically relevant pathotypes of ExPEC, that are causal agents of a spectrum of diseases in mammals from uncomplicated urinary tract infections (UTIs), to bacteraemia, and meningitis. ExPEC are distinct amongst *E. coli* in their ability to colonise several anatomical niches within the host. ExPEC strains are able to survive outside of the intestine, partly due to their ability to catabolise a diverse range of host metabolites, including D-serine. UPEC and NMEC carry a locus, *dsdCXA*, for D-serine metabolism. DsdC, a LysR-type transcriptional regulator (LTTR), autoregulates its own expression and further controls the expression of DsdX, a D-serine inner membrane transporter and DsdA, a D-serine deaminase that catabolises D-serine into pyruvate and ammonia. Strikingly, NMEC strains carry two copies of the *dsdCXA* locus: *dsdCXA1* and *dsdCXA2*.

Metabolism and virulence are often interlinked in bacteria, with host metabolites, in addition to being carbon sources, acting as environmental signals that can affect bacterial gene expression. It was thus hypothesised that D-serine, which is abundant in the brain and urinary tract, may act as an environmental stimulus that enables two ExPEC pathogens to sense their environment and modify their gene expression, through the regulatory actions of DsdC.

In this work, using a combination of chromatin immunoprecipitation coupled with next generation sequencing (ChIP-Seq) and RNA sequencing (RNA-Seq), the direct regulon of DsdC in two prototypical UPEC and NMEC strains, CFT073 and CE10, was elucidated, in the presence and absence of D-serine. It was shown that DsdC bound to distinct regions of both the CFT073 and CE10 genome, indicating that DsdC has been tailored for strain-specific lifestyles. In CFT073, it was shown that DsdC bound and affected expression of genes involved in colibactin synthesis, a bacterial genotoxin. Furthermore, it was revealed that in CE10, DsdC bound and affected expression of several genes involved in outer membrane-associated virulence, such as capsular and LPS biosynthesis genes. Using a range of biochemical techniques, it was established that DsdC, upon exposure to D-serine, modified the O-antigen structure of CE10. Furthermore, it was revealed that DsdC1, but not DsdC2, affected the survival of CE10 against K1-specific bacteriophages, through the regulation of the *O*-acetyltransferase gene *neuO*, elucidating a distinct role for the homologous proteins. Collectively, this work describes how regulation by DsdC has been tailored to suit strain-specific lifestyles in two ExPEC strains and further enhances our understanding of how bacterial pathogens sense their environment and mediate pathogenesis.

Author’s declaration

I hereby declare that this thesis is the result of my own work and has been composed for the degree of PhD at the University of Glasgow. This work has not been submitted for any other degree at this or any other institution. All work presented was performed by myself unless otherwise stated. All sources of information and contributions to the work have been specifically acknowledged in the text.

Natasha C A Turner

August 2021

Chapter 1 Introduction

1.1 *Escherichia coli*

Escherichia coli is a Gram-negative, facultative anaerobe, first described by Theodor Escherich in the late nineteenth century (Hacker & Blum-Oehler, 2007). For decades, it has been used as a model organism for bacterial studies, often being considered as the molecular biology 'workhorse' of the laboratory (Hacker & Blum-Oehler, 2007). *E. coli* is both a commensal and pathogenic organism, commonly found within the microbiota of mammals and birds (Rasko *et al.*, 2008). The composition of the microbiome is dynamic, however, *E. coli* is one of 109 species that form the core baseline microbiome in healthy controls (King *et al.*, 2019). Commensal *E. coli* are typically found within the caecum and colon of the large intestine in mammals, residing in the mucus layer covering the epithelial cells (Li *et al.*, 2015; Poulsen *et al.*, 1994; Tenaillon *et al.*, 2010). Typically, *E. coli* is one of the first bacterial species that colonises neonates (Casaburi *et al.*, 2021), providing benefits to the host including preventing the colonisation of other pathogens (reviewed in Tenaillon *et al.*, 2010). Conversely, through both intestinal and extraintestinal disease, *E. coli* is thought to kill more than 2 million humans per year (reviewed in Tenaillon *et al.*, 2010).

E. coli is broadly divided into eight main phylogenetic groups: A, B1, B2, C, D, E, F, and G (Denamur *et al.*, 2021). An additional group closely related to D, termed H, has also been suggested (Lu *et al.*, 2016). Analysis of multiple *E. coli* genomes has revealed large differences in size between strains, varying from 4.2 to 6 Mbp, corresponding to 3,900 to 5,800 genes respectively (reviewed in Denamur *et al.*, 2021). This genetic diversity can be ascribed to the relative plasticity of the *E. coli* genome (Touchon *et al.*, 2009). Indeed, through horizontal gene transfer (HGT), genes can be acquired laterally, on mobile genetic elements (MGE) such as transposons, plasmids, bacteriophages or genomic islands (reviewed in Juhas, 2015). Studies have suggested that up to 75% of genes in each genome have been acquired through HGT (reviewed in Juhas, 2015). Many of the characteristic virulence factors carried by *E. coli* are encoded on pathogenicity-associated islands (PAI) and are inserted into the genome in tRNA 'hotspots' (Johnson & Russo, 2002). The genome size of *E. coli* is finite however, with multiple selective pressures often dictating acquisition or loss of genetic elements. Earlier studies reported that all *E. coli* shared a core genome of ~2,000 genes, with the remaining genes obtained from a pangenome

of ~18,000 (Touchon *et al.*, 2009). However, ~26 new genes are purported to be identified with each new strain sequenced, leading to the suggestion that the pangenome now consists of ~75,000 genes (reviewed in Denamur *et al.*, 2021). This would indicate that the *E. coli* genome is potentially inexhaustible.

Through HGT and evolution, several pathotypes of *E. coli* have emerged across the phylogroups. Based upon clinical data, *E. coli* is classified into three main groups: commensal strains; intestinal pathogenic *E. coli* (InPEC); and extraintestinal pathogenic *E. coli* (ExPEC) (Russo & Johnson, 2000). Within the InPEC group, there are six main *E. coli* pathotypes recognised: Shiga-toxin producing/enterohaemorrhagic *E. coli* (STEC/EHEC); enteropathogenic *E. coli* (EPEC); enteroinvasive *E. coli* (EIEC); enteroaggregative *E. coli* (EAEC); enterotoxigenic *E. coli* (ETEC); and diffusely adherent *E. coli* (DAEC) (Russo & Johnson, 2000). These pathotypes cause enteric disease and, although there is some overlap, they each have a distinct repertoire of virulence factors (Russo & Johnson, 2000).

Conversely, whilst InPEC pathogenesis is consigned to the intestinal tract, ExPEC strains can infect almost any organ or anatomical region within the host (Russo & Johnson, 2000). There are several pathotypes that comprise ExPEC strains: neonatal meningitis-associated *E. coli* (NMEC); uropathogenic *E. coli* (UPEC); and sepsis-associated *E. coli* (SEPEC). ExPEC strains predominantly reside asymptomatically within the intestinal tract, causing disease only upon egression of the gut (Johnson & Russo, 2002). ExPEC strains often employ a range of virulence factors such as adhesins, toxins, and host avoidance systems, to help them survive in extraintestinal niches (Kaper *et al.*, 2004).

1.2 Extraintestinal pathogenic *E. coli* (ExPEC)

ExPEC isolates are able to infect multiple anatomical niches within the host, causing urinary tract, bloodstream, prostate, and further infections (Manges *et al.*, 2019). ExPEC isolates have been defined as carrying at least two of the following virulence factors: P-fimbriae (*pap*); S-fimbriae/F1C-fimbriae (*sfa*/*foc*); Dr binding proteins (*afa*/*draBC*); capsule (*kpsM II);* or the aerobactin receptor (*iutA*) (Dale & Woodford, 2015). Indeed, it is the collection of virulence factors that enables ExPEC strains to cause disease, with no single virulence factor

acting as the causal agent (Dale & Woodford, 2015). The virulence factor profiles of ExPEC isolates are therefore highly variable; however, as with all *E. coli*, ExPEC strains have an outer membrane layer comprising lipopolysaccharides (LPS) among other components. *E. coli* LPS plays a key role in virulence and protection of the cell from environmental and immune system insults (Ebbensgaard *et al.*, 2018).

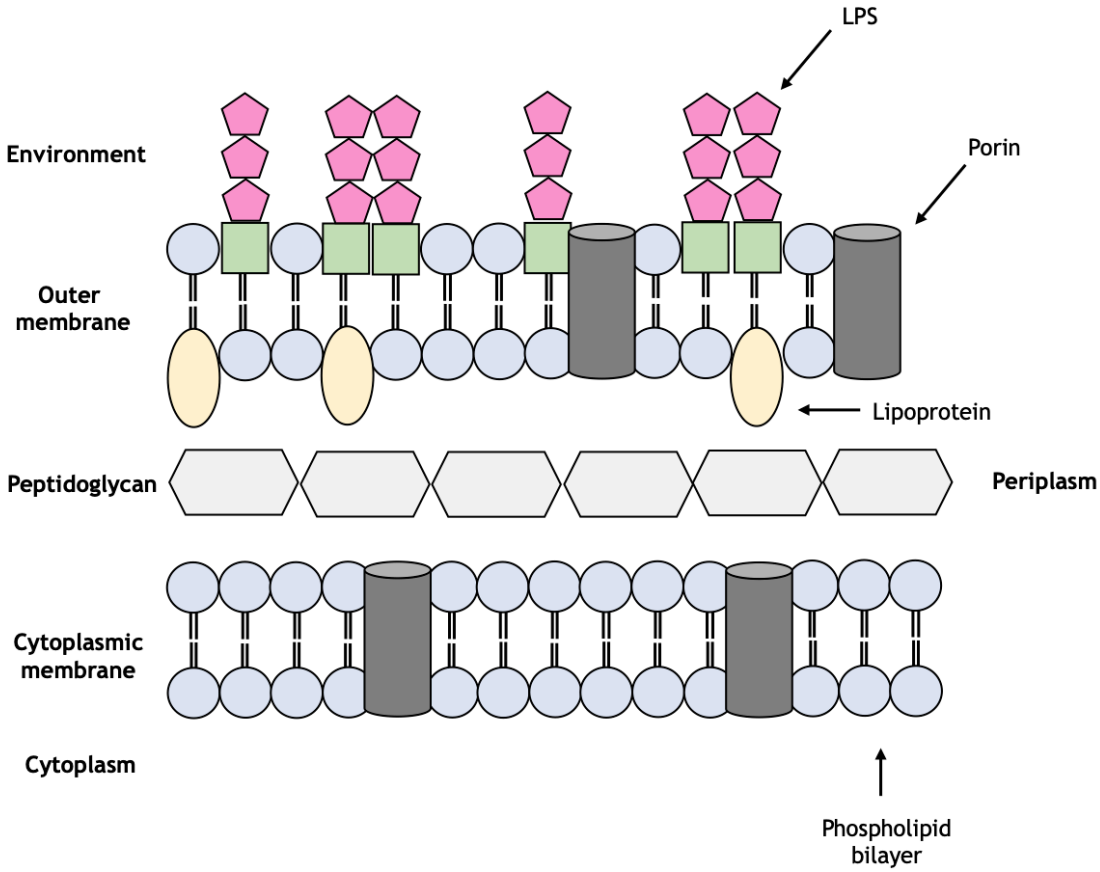


Figure 1-1 Composition of the Gram-negative cell wall. Adapted from Schwechheimer & Kuehn, 2015. A Gram-negative cell envelope is composed of three elements: an outer membrane; peptidoglycan; and an inner membrane. The inner leaflet of the outer membrane is composed of phospholipids, and the outer leaflet is composed of LPS and envelope proteins (porins). The cytoplasmic membrane (or inner membrane) is composed of a phospholipid bilayer. Between the two membranes is the periplasm, which contains the peptidoglycan layer.

LPS is composed of three components: lipid A, a hydrophobic protein that embeds the LPS structure into the outer membrane; the core polysaccharide which is comprised of multiple sugars; and the hypervariable O-antigen, which comprises numerous repeating sugar units (reviewed in Liu *et al.*, 2020). In non-

encapsulated *E. coli*, the LPS is exposed to the host environment and is often the first line of defence against the host (reviewed in Bertani & Ruiz, 2018). Fundamentally, LPS is a structural component of the outer membrane, and functions as a permeability barrier to the environment (reviewed in Bertani & Ruiz, 2018). However, LPS is also an endotoxin, with lipid A activating the host immune response and the inflammatory system (reviewed in Raetz & Whitfield, 2002). The host's immune response is dependent on the concentration of LPS, with high concentrations of LPS inducing fever and increasing the heart rate within the host, which can lead to septic shock and death (reviewed in Wang & Quinn, 2010). Alternatively, due to the hypervariability of the O-antigen, *E. coli* can evade the immune response and so is protected from host-mediated lysis and phagocytosis (reviewed in Bertani & Ruiz, 2018). The LPS therefore plays a key role in protection and virulence in ExPEC, and indeed all *E. coli*, strains.

1.2.1 Neonatal meningitis-associated *E. coli* (NMEC) prevalence and pathogenesis

Neonatal meningitis is a severe disease that contributes to neonatal morbidity and mortality worldwide, and is characterised as the inflammation of the meninges during the first 28 days of life (Khalessi & Afsharkas, 2014). The incidence of bacterial-associated neonatal meningitis ranges from 0.25-1 per 1,000 live births (Khalessi & Afsharkas, 2014). However, it has been reported that the incidence of meningitis is much greater in lower-income countries, where it can be as high as 0.8-6.1 per 1,000 live births (reviewed in Ku, Boggess & Cohen-Wolkowiez, 2015). The mortality rate in lower-income countries is ~40%, with some 50,000 new-borns dying every year, thereby making meningitis one of the five most prevalent neonatal infections worldwide (reviewed in Bonacorsi & Bingen, 2005). For the neonates that survive, neurological sequelae can occur in 30-50% of cases, presented as learning difficulties, blindness, hydrocephalus, and hearing loss (reviewed in Mann & Jackson, 2008).

E. coli is the second largest etiological agent of neonatal meningitis, after Group B *Streptococcus* (reviewed in Bonacorsi & Bingen, 2005). Indeed, NMEC is the cause of 20-40% of the cases that occur in the USA annually (reviewed in Russo & Johnson, 2003). The most common way in which meningitis develops in neonates is by a primary blood stream infection. Initially, NMEC is acquired in the neonate

from the mother's flora, or from environmental sources (reviewed in Bonacorsi & Bingen, 2005). NMEC then establishes itself within the intestinal tract, before translocating to the bloodstream (reviewed in Bonacorsi & Bingen, 2005). Once there is sufficient bacteraemia (10^3 colony forming units (CFU)/ml) within the bloodstream, NMEC penetrates the blood-brain barrier (BBB), invading the central nervous system (CNS) (Huang *et al.*, 1995; reviewed in Xie, Kim & Kim, 2004). Once NMEC passes through the BBB and enters the cerebral spinal fluid (CSF), it is liberated from the immune system and is therefore able to proliferate freely (reviewed in Polin & Harris, 2001). Natural bacterial lysis subsequently causes release of bacterial cell wall products, which act as stimuli for the release of host inflammatory cytokines, and it is this response that causes meningitis (reviewed in Polin & Harris, 2001).

Although the mechanism has not been fully elucidated, a key step in NMEC pathogenesis is traversal of the BBB. The BBB is a structural barrier that is formed by human brain microvascular endothelial cells (HBMEC), pericytes and astrocytes, and which protects the brain from a number of insults (reviewed in Kim, 2012). NMEC has been shown to attach to the BBB using several factors including FimH of the type I pili, which binds to the host receptor CD48, and outer membrane protein A (OmpA) which binds to the host receptor ECGP96 (Khan *et al.*, 2007; Prasadarao, 2002; Prasadarao *et al.*, 1996). Further, transposon mutagenesis revealed that NMEC invasion also occurs through the actions of proteins that are termed the invasion of brain endothelial cell (Ibe) adhesins, although the receptors for this are unknown (Huang *et al.*, 1995, 1999). These interactions trigger a cascade of host cell signalling pathways, resulting in an increase in intracellular Ca^{2+}, which stimulates actin cytoskeleton rearrangements, a prerequisite for NMEC invasion (Khan *et al.*, 2007; Maruvada & Kim, 2012). Cytotoxic necrotising factor 1 (CNF1) also contributes to NMEC invasion of the BBB, and has been shown to play a role in host cell myosin rearrangement, as well as actin rearrangement (Khan *et al.*, 2002; Essler *et al.*, 2003; reviewed in Croxen & Finlay, 2010). CNF1 binds to the host receptor 67LR, inducing cytoskeleton rearrangements, promoting internalisation of NMEC into HBMECs (Kim *et al.*, 2005). These data reveal the multifactorial processes that NMEC can use in order to attach and traverse the BBB.

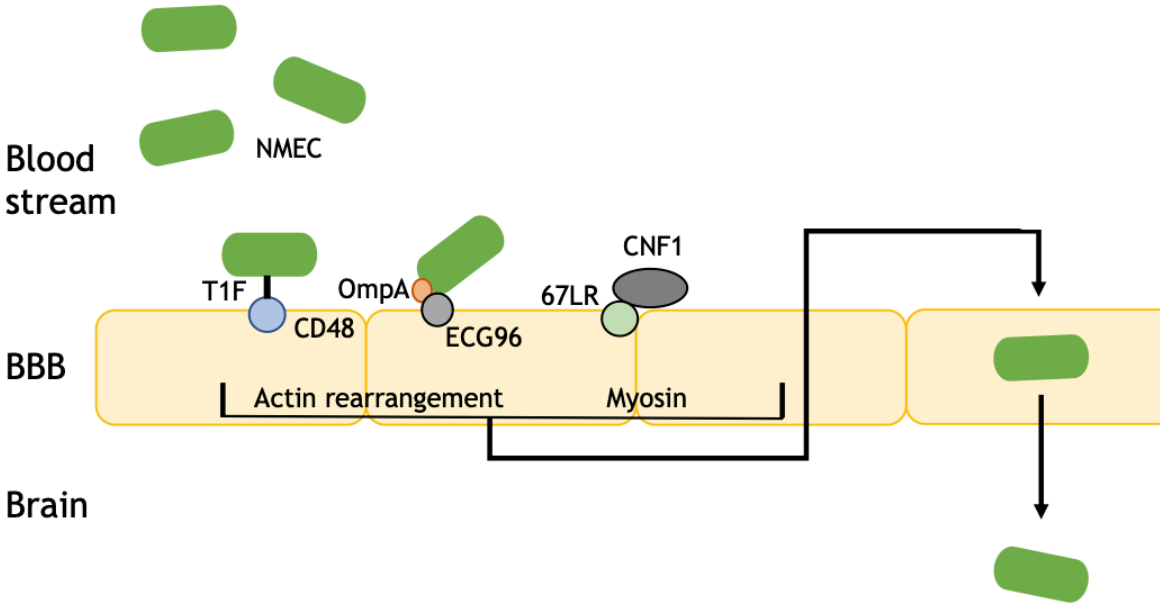


Figure 1-2 Mechanisms of NMEC invasion of the BBB. Adapted from Croxen & Finlay, 2010. Attachment of NMEC to the BBB is mediated by type 1 pili binding to the host receptor CD48 and OmpA binding to the host receptor ECG96. These interactions mediate an increase in intracellular Ca^{2+} which stimulates actin rearrangements. Further, release of CNF1 mediates binding to host receptor 67LR which causes myosin rearrangements, enabling NMEC to invade the brain endothelial cells.

1.2.2 NMEC K1 polysialic acid capsule

Survival within the blood is crucial for eliciting pathogenesis at the BBB interface, as high bacteraemia is required (Huang *et al.*, 1995). NMEC strains therefore characteristically have an antiphagocytic capsule that surrounds the cell and provides protection from the host immune response (reviewed in Croxen & Finlay, 2010). There are over 70 distinct capsular (K) antigens found in *E. coli* (Jann & Jann, 1992), however, the K1 antigen is often over-represented in NMEC infection, with a study showing that 84% of isolates recovered from the CSF of neonates with meningitis had the K1 capsular polysaccharide (Robbins *et al.*, 1974). Furthermore, mutants devoid of the K1 capsule were unable to be detected in the CSF in an *in vivo* rat model, indicating the importance of the K1 capsule to infection (Kim *et al.*, 1992).

A key survival mechanism of bacteria is the ability to evade the host immune response. Indeed, as discussed above, there are over 70 distinct capsular (K)

antigens that have been found in *E. coli* (Jann & Jann, 1992), with each K antigen differing in its sugar composition, thereby presenting the host immune system with a range of differing structures. Moreover, the K1 polysaccharide is homopolymer of α2,8-linked sialic acid residues, which in itself is a molecular mimic of the host polysialic acid of the neural cell adhesion molecule (N-CAM) (reviewed in Silver & Vimr, 1990). K1 polysaccharides are therefore weakly immunogenic in mammals and do not stimulate synthesis of antibodies (reviewed in Silver & Vimr, 1990).

The mammalian complement system also plays a crucial role in defending the host against pathogens and represents one of the main mechanisms of innate immunity (Dunkelberger & Song, 2010). The complement system can be activated through three major pathways: classical, lectin and alternative (Dunkelberger & Song, 2010). The K1 capsule has been reported to play a role in the ability of NMEC to evade complement-mediated killing, with strains that lose the ability to synthesise K1 subsequently losing serum resistance properties (Cross *et al.*, 1986; Leying *et al.*, 1990). Factor H, the main soluble regulator of the alternative complement pathway, has been reported to bind to polyanions, such as the K1 polysialic acid capsule (reviewed in Abreu & Barbosa, 2017). This interaction has been suggested to downregulate the alternative complement pathway, thus impairing the membrane attack complex (reviewed in Abreu & Barbosa, 2017). Moreover, the presence of an extracellular polysaccharide, such as the K1 capsule, is thought to mask underlying bacterial cell surface structures that could activate the complement system (reviewed in Silver & Vimr, 1990). The K1 capsular polysaccharide therefore plays a key role in the virulence of NMEC isolates, enabling them to evade and resist the host immune system.

1.2.3 Further NMEC-associated virulence factors

As well as the K1 capsule, NMEC strains use a wide range of virulence factors in order to survive and evade the host. These factors include iron acquisition systems, invasion proteins (*ibe*), toxins (CNF1), and a cryptic type three secretion system (T3SS), termed the *Escherichia coli* type III secretion system 2 (ETT2) (Lu *et al.*, 2011; Wijetunge *et al.*, 2015). CNF1 is a protein toxin that can be produced by NMEC, and which induces the development of stress fibres in epithelial cells via the Rho pathway (Lemichez *et al.*, 1997). An early study

suggested that CNF1 induced high lethality when injected into mice and caused the formation of multi-nucleated cells *in vitro* (Lemichez *et al.*, 1997). Furthermore, CNF1 has been reported to activate NF-kB in epithelial cells, thereby stimulating the expression of pro-inflammatory factors (reviewed in Fabbri, Travaglione & Fiorentini, 2010). Conversely, further research reported that injection of the entire CNF1 toxin caused no lethal effects in animals, and reported that CNF1 might instead play a role in protecting the host cell from apoptotic stimuli (reviewed in Fabbri, Travaglione & Fiorentini, 2010). Although this has not been proven, this protection may facilitate NMEC survival within the niche by keeping the host cell alive. However, CNF1 plays a key role in invasion of NMEC into the CNS (Khan *et al.*, 2002), indicating that although its role in the host has not been fully elucidated, it is an important virulence determinant for NMEC.

T3SS are macromolecular protein syringes that are unique to Gram-negative bacteria, and are used to inject virulence factors into host cells or bacterial competitors (reviewed in Slater *et al.*, 2018). The T3SS spans the inner and outer membranes of Gram-negative bacteria, with a needle-like filament protruding into the environment. In *E. coli*, the T3SS is best characterised in EHEC and EPEC strains, which can use the T3SS to form attaching and effacing lesions on host cells and deliver effector proteins (Mellies *et al.*, 1999; Moon *et al.*, 1983). In EHEC and EPEC, the genes for the T3SS are carried on a pathogenicity-associated island (PAI) termed the locus of enterocyte effacement (LEE) (McDaniel *et al.*, 1995). Intriguingly, some NMEC strains have been reported to carry the apparatus genes of a cryptic secondary T3SS, ETT2 (Yao *et al.*, 2009). This secondary T3SS was originally first described in EHEC, having undergone widespread mutational attrition, rendering it unable to form a functional secretion system (Ren *et al.*, 2004). However, studies have reported that the ETT2 plays a role in virulence in bloodstream isolates of *E. coli*, with mutations in CE10 of the ETT2 leading to defects in invasion and intracellular survival in HBMECs (Yao *et al.*, 2009). Although the role of the ETT2 has not been fully elucidated, this data does indicate that the ETT2 is an important virulence factor in NMEC.

1.2.4 Uropathogenic *E. coli* (UPEC) prevalence and pathogenesis

Urinary tract infections (UTIs) are a common disease, and are estimated to affect up to 150 million people per year, posing a significant economic and health burden upon society (Flores-Mireles *et al.*, 2015). Indeed, it is estimated that the societal cost of UTIs in the United States alone is some $3.5 billion per year (Flores-Mireles *et al.*, 2015). UTIs are often referred to as being either uncomplicated or complicated infections. Uncomplicated UTIs are classified as lower UTIs, which affect the urinary tract/bladder (cystitis), or as upper UTIs, which affect the upper urinary tract/kidneys (pyelonephritis) (Mazzulli, 2012). These often occur in otherwise healthy individuals and can cause symptoms of dysuria and fever. Complicated UTIs often develop in individuals with underlying illnesses or who are immunocompromised, amongst other factors (Mazzulli, 2012). Complicated UTIs can lead to severe pyelonephritis, renal scarring and end-stage disease (Melekos & Naber, 2000). UTIs are caused by a range of different pathogenic organisms, including *Klebsiella pneumoniae*, *Proteus mirabilis* and *Staphylococcus saprophyticus* (Mazzulli, 2012). However, the most common infectious agent of UTIs is UPEC. Indeed, 75-95% of uncomplicated UTIs are associated with UPEC strains (Hooton, 2012).

The urinary tract was previously classed as a sterile environment, however, metagenomic analysis of the urinary tract has revealed a wealth of information, indicating the existence of a urogenital microbiome (Thomas-White *et al.*, 2018). UPEC pathogenesis occurs when UPEC strains translocate from the gut and colonise the periurethral and urethra (reviewed in Flores-Mireles *et al.*, 2015). UPEC can then ascend into the bladder lumen and adhere, using type 1 fimbriae (T1F) which bind to uroplakin receptors on bladder epithelial cells (Connell *et al.*, 1996; Mulvey *et al.*, 1998). This binding, and further interactions, stimulates invasion of the epithelial cells and triggers local actin rearrangement, resulting in internalisation of the attached UPEC (Martinez *et al.*, 2000). UPEC can then proliferate within the host cell and biofilm-like complexes are formed, termed intracellular bacterial communities (IBCs) (Anderson *et al.*, 2003). The formation of IBCs protects UPEC from neutrophils, antibiotics and other environmental insults (Anderson *et al.*, 2003). Subsequent UPEC invasion and attachment results in an influx of polymorphonuclear neutrophils (PMNs), which cause tissue

damage and apoptosis of the bladder cells (Mulvey *et al.*, 1998; Haraoka *et al.*, 1999; reviewed in Croxen & Finlay, 2010). Left untreated, some UPEC strains can disseminate to the kidney, causing pyelonephritis, and further, penetrate through the endothelial cells into the bloodstream, resulting in bacteraemia (reviewed in Kaper, Nataro & Mobley, 2004) (Fig 1-3).

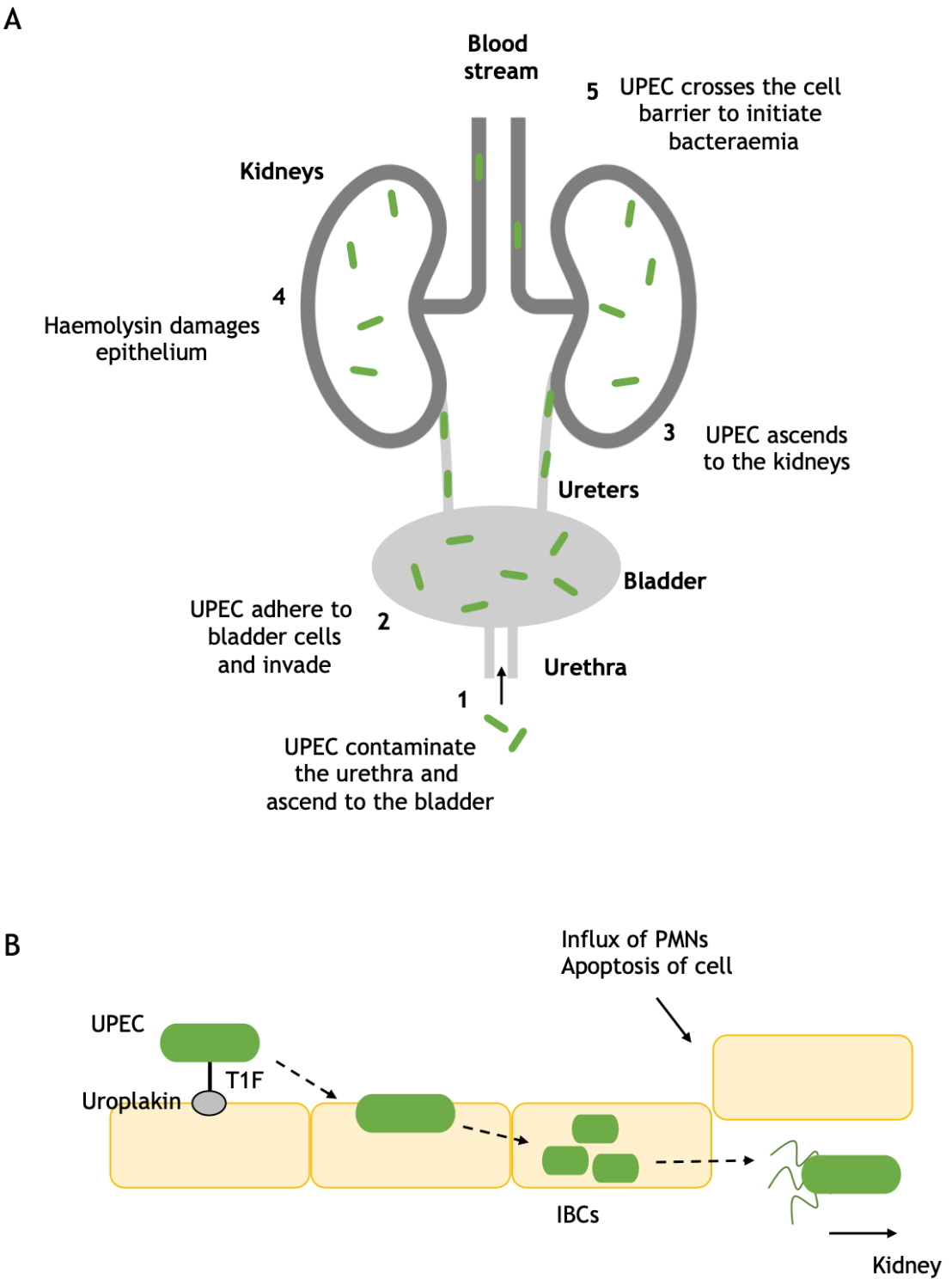


Figure 1-3 Mechanism of UPEC pathogenesis progression within the host. (A) Adapted from Kaper, Nataro & Mobley, 2004. (1) UPEC contaminate the urethra

and ascend to the bladder. (2) UPEC bind to the epithelial cells of the bladder and invade. (3) Some UPEC ascend up the ureters to the kidney. (4) After disseminating to the kidney, UPEC can release haemolysins, damaging host epithelial cells. (5) UPEC can disseminate further, crossing the cell barrier into the blood supply to initiate bacteraemia. (B) Adapted from Croxen & Finlay, 2010. UPEC attach to the uroplakin receptor on the bladder epithelial cells using T1F. This binding stimulates an unknown pathway which mediates invasion into the host cell. UPEC then form IBCs within the host cell. These interactions cause an influx of PMNs. Apoptosis of the bladder cells occurs. UPEC strains can disseminate from the bladder to the kidneys.

1.2.5 UPEC-associated virulence factors

There are four main phylogenetic groups in which UPEC strains have been identified; A, B1, B2, and D, highlighting the diversity of the pathotype (Terlizzi *et al.*, 2017). Indeed, there is not a single set of virulence determinants that defines the UPEC pathotype. UPEC strains possess numerous virulence factors that enable them to cause disease in hosts. These include structural mechanisms such as fimbriae, pili, and flagella, as well as secreted mechanisms such as outer membrane proteins (OMPs), iron-acquisition systems and bacterial toxins (Najafi *et al.*, 2018; Shah *et al.*, 2019; Welch *et al.*, 2002).

UPEC secrete several bacterial toxins, including colibactin and α-haemolysin (Welch *et al.*, 2002). α-haemolysin is encoded on the *hlyCABD* operon in *E. coli* and is a member of the repeats-in-toxin (RTX) family of toxins (Felmlee, Pellett & Welch, 1985; reviewed in Welch, 2016). *In vitro* studies have revealed that α-haemolysin is cytotoxic to renal cells (Mobley *et al.*, 1990). Indeed, α-haemolysin has been reported to oligomerise and integrate into the host cell membrane, resulting in pore formation and subsequent lysis of the host cell (Bhakdi & Tranum-Jensen, 1986; Benz *et al.*, 1989; reviewed in Flores-Mireles *et al.*, 2015). This mechanism has been reported to facilitate the acquisition of iron and nutrients for the bacteria (reviewed in Flores-Mireles *et al.*, 2015). Further, α-haemolysin has been reported to be involved in pathways that could trigger exfoliation of the epithelial cells, exposing deeper layers of host cells for further UPEC colonisation (Dhakal & Mulvey, 2012).

A further important virulence factor for UPEC are iron-acquisition systems. Iron is limited within the urinary tract, however it is essential for bacterial growth (Watts *et al.*, 2012). Bacteria therefore utilise siderophore systems for scavenging iron. Siderophore systems comprise of siderophores, which are low-molecular-weight Fe^{3+}-chelating compounds, and a membrane receptor, which internalises the siderophore-bound iron (reviewed in Flores-Mireles *et al.*, 2015). In *E. coli*, four different siderophore systems have been identified (Watts *et al.*, 2012), indicating the importance of iron-acquisition to UPEC. Two siderophores, aerobactin and yersiniabactin, have been reported to be the most critical in iron acquisition in the urinary tract (Garcia *et al.*, 2011). Further, genomic studies *in vivo* using UPEC strain UTI89 revealed, relative to samples obtained from the caecum, IBCs showed a marked increase in expression of components of iron-acquisition systems (Reigstad *et al.*, 2007), again indicating the importance of iron-acquisition in UPEC pathogenesis.

1.3 Bacterial transcriptional regulation

Fine-tuning of gene expression in response to environmental cues is a key survival mechanism for bacteria. Gene expression occurs when template DNA is transcribed into mRNA by RNA polymerase, followed by translation of the mRNA into a functional protein by ribosomal machinery. Transcription can be split into three main steps: initiation, elongation, and termination. Transcriptional regulation can occur at any stage within the transcription process, but most frequently occurs within the initiation stage (reviewed in Browning & Busby, 2004).

1.3.1 Initiation of transcription

All bacterial transcriptional initiation is dependent on RNA polymerase (RNAP). The core enzyme of RNAP is comprised of five subunits: two large β subunits, β and β'; two identical α subunits; and a ω subunit. The two α subunits consist of two domains, an α-N-terminal domain (αNTD) and an α-C-terminal domain (αCTD). The αNTD is responsible for the assembly of the β subunits, which is the active site of the polymerase, and the αCTD plays a role in regulating transcription and binds to DNA (reviewed in Browning & Busby, 2004; Murakami, 2015). The ω subunit of RNAP had long remained elusive to scientists; however,

research has revealed that it plays a structural role in maintenance of β' and further acts as a chaperone of β' to the enzyme assembly (Fig. 1-4A) (reviewed in Mathew & Chatterji, 2006).

Although RNAP is required for transcription, it can only weakly bind in an indiscriminate fashion to promoter DNA (reviewed in Ghosh, Bose & Zhang, 2010). Directed transcription cannot be initiated without an additional protein, a σ factor, which is used to recognise specific promoter sequences. Promoter sequences consist of several unique elements, including the -10 and -35 consensus sequence, and it is these two sequence motifs that are recognised by the σ factor. The combined complex of RNAP and σ factor is called the holoenzyme, and it is this complex that can bind to DNA and initiate targeted transcription (Fig.1-4B) (reviewed in Browning & Busby, 2016). Once the σ factor has recruited the RNAP and formed a holoenzyme, it interacts with the promoter region, forming a closed complex (RPc). This closed complex can then switch to an open complex (RPo), which involves the unwinding of ~12 bp of double stranded DNA (dsDNA). The template strand of DNA is then loaded into the RNAP active site for RNA synthesis. The σ factor is released and RNAP enters the elongation phase (Fig.1-4C). Many factors can influence the initiation of transcription including σ factors, transcription factors (TFs), ligands, and nucleoid associated proteins (NAPs), together enabling fine-tuning of bacterial gene expression.

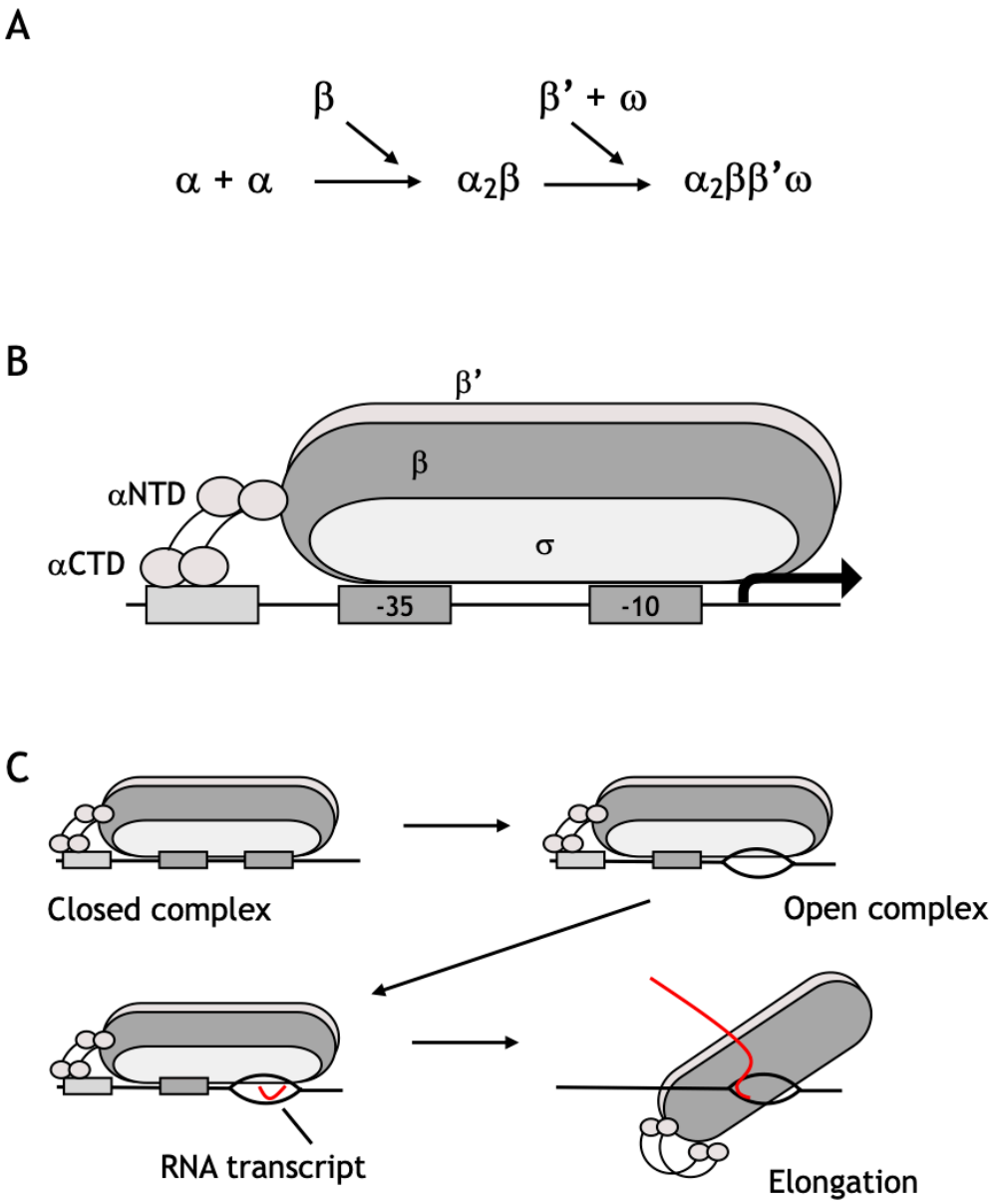


Figure 1-4 Initiation of transcription by bacterial RNA polymerase. (A) General structure of RNAP assembly, adapted from Murakami, 2015. (B) A graphical representation of the RNAP holoenzyme, adapted from Browning & Busby, 2004. The two β subunits, β and β', form the catalytic part of the RNAP, assembled together by αNTD. The αCTD binds to the promoter region DNA, upstream of the -35 and -10 elements. The σ factor binds to the -10 and -35 elements, initiating transcription. (C) A graphical representation of transcription initiation, adapted from Browning and Busby, 2016. The RNAP holoenzyme binds to the promoter region, forming a closed complex. The closed complex can then switch to an open complex and the dsDNA is unwound. The template strand of DNA is loaded into the active site and RNA transcript is formed. The σ factor is then released and RNAP enters the elongation phase.

1.3.2 Sigma (σ) factors

As discussed above, σ factors are used to recognise specific sequences of promoter regions, directing the RNAP to the appropriate transcriptional start sites (TSS) and initiating strand separation of DNA to form the transcription bubble. Bacteria have multiple σ factors, with *E. coli* carrying seven: σ^{70}, σ^{54} (σ^{N}), σ^{FecI}, σ^{24} (σ^{E}), σ^{28} (σ^{F}), σ^{32} (σ^{H}), σ^{38} (σ^{S}). σ^{70} is the essential housekeeping σ factor, which regulates the transcription of thousands of genes. Conversely, the other six σ factors are referred to as the alternative σ factors, regulating genes involved in numerous stress responses as detailed in Table 1-1 (reviewed in Ishihama, 2000). All *E. coli* σ factors belong to the σ^{70} family, except σ^{54} which is evolutionarily unrelated to the other σ factors (Lonetto, Gribskov & Gross, 1992; reviewed in Merrick, 1993).

Sigma (σ) factor	Sigma (σ) factor regulon
σ^{70}	Housekeeping
σ^{54} (σ^{N})	Nitrogen stress response
σ^{FecI}	Ferric citrate transport
σ^{24} (σ^{E})	Heat shock response/cell envelope stress
σ^{28} (σ^{F})	Flagella/chemotaxis response
σ^{32} (σ^{H})	Heat shock response
σ^{38} (σ^{S})	Stationary phase/general stress response

Table 1-1 *E. coli* σ factors and their respective regulons. A table of the seven *E. coli* σ factors and the functions they regulate.

The concentration of each σ factor is dependent on cell growth conditions: σ^{70} is reported to be the most abundant, and has the highest affinity for RNAP (reviewed in Ishihama, 2000; Paget, 2015). Each σ factor is composed of multiple domains, with σ^{70} comprised of four domains: $\sigma_{1.1}$, σ_2, σ_3, and σ_4. Each domain has DNA binding properties, except for $\sigma_{1.1}$, which has been reported to have a role in allowing access to the RNAP active site. The other domains bind to the promoter region: σ_2 binds to the -10 element, σ_3 binds to the extended -10 element, and σ_4 binds to the -35 element (reviewed in Feklístov *et al.*, 2014). As the different σ factors guide the RNAP to different promoter regions depending on the growth conditions and stresses, σ factors represent a simple mechanism of gene regulation in bacteria. The activity of σ factors, and subsequent directed bacterial gene expression, is further controlled by anti-σ factors. Anti-σ

factors can sequester σ factors, impeding RNAP binding. Indeed, in *E. coli* stationary phase growth, Rsd interacts with σ^{70} domains σ_2 and σ_4, preventing binding to the RNAP and the -35 element, allowing σ^{38} to bind to RNAP, promoting expression of stationary phase related genes (reviewed in Paget, 2015). Gene regulation by σ factors is therefore both a simple and complex mechanism of gene regulation in bacteria. Conversely, transcription often cannot occur without the presence of a further protein or ligand, such as TFs.

1.3.3 Transcription factors (TFs), nucleoid associated proteins (NAPs), and two-component systems (TCS)

Both TFs and NAPs are DNA-binding proteins that can influence gene regulation. Around 300 genes in *E. coli* have been predicted to be DNA-binding proteins, indicating the importance of these regulatory elements to the cell (Babu & Teichmann, 2003a; Martínez-Antonio & Collado-Vides, 2003). However, it has been estimated that just seven of these TFs control up to 50% of all regulated genes (Martínez-Antonio & Collado-Vides, 2003). TFs are often expressed in response to host environmental stimuli, with some TFs auto-regulating themselves (reviewed in Browning & Busby, 2004; Maddocks & Oyston, 2008).

In order to regulate gene expression, TFs are reliant on recognising DNA motifs at promoter regions (D'haeseleer, 2006; Goethals *et al.*, 1992). Most TFs therefore carry a DNA-binding motif, although this is usually insufficient to facilitate binding, and thus most TFs bind as dimers (reviewed in Browning, Butala & Busby, 2019). TFs activate or repress the expression of genes by several mechanisms, usually depending on environmental signals. There are two general mechanisms for activation of genes by TFs: simple activation and activation by conformational change (Fig. 1-5). In simple activation, the TF binds upstream of the promoter element and recruits RNAP by directly interacting with either the σ_4 element or the αCTD (Zhou, Zhang & Ebright, 1993; Busby & Ebright, 1994). TFs can also positively regulate gene expression by binding to, or near, the promoter element and altering the conformation, realigning the -10 and -35 element, and thus enabling RNAP to bind (Heldwein & Brennan, 2001).

TFs can also repress gene expression. There are three general mechanisms through which this can occur: steric hinderance, DNA looping, and locking RNAP

at the promoter (Fig. 1-5). The simplest form of gene repression occurs when repressors bind within the promoter element, thereby causing steric hinderance and occluding RNAP from the site (Jacob & Monod, 1961). Further, to repress gene expression, singular or multiple TFs may bind to promoter-distal sites and thus DNA looping may occur, thereby hindering RNAP from binding to the promoter element (Vörös *et al.*, 2017). Finally, TFs can block gene expression by locking RNAP at the promoter region and thus transcription cannot proceed (Grainger *et al.*, 2008). Although these are the main mechanisms of TF regulation, there are numerous other mechanisms that can involve multiple TFs at complex promoter regions (Browning, Butala & Busby, 2019).

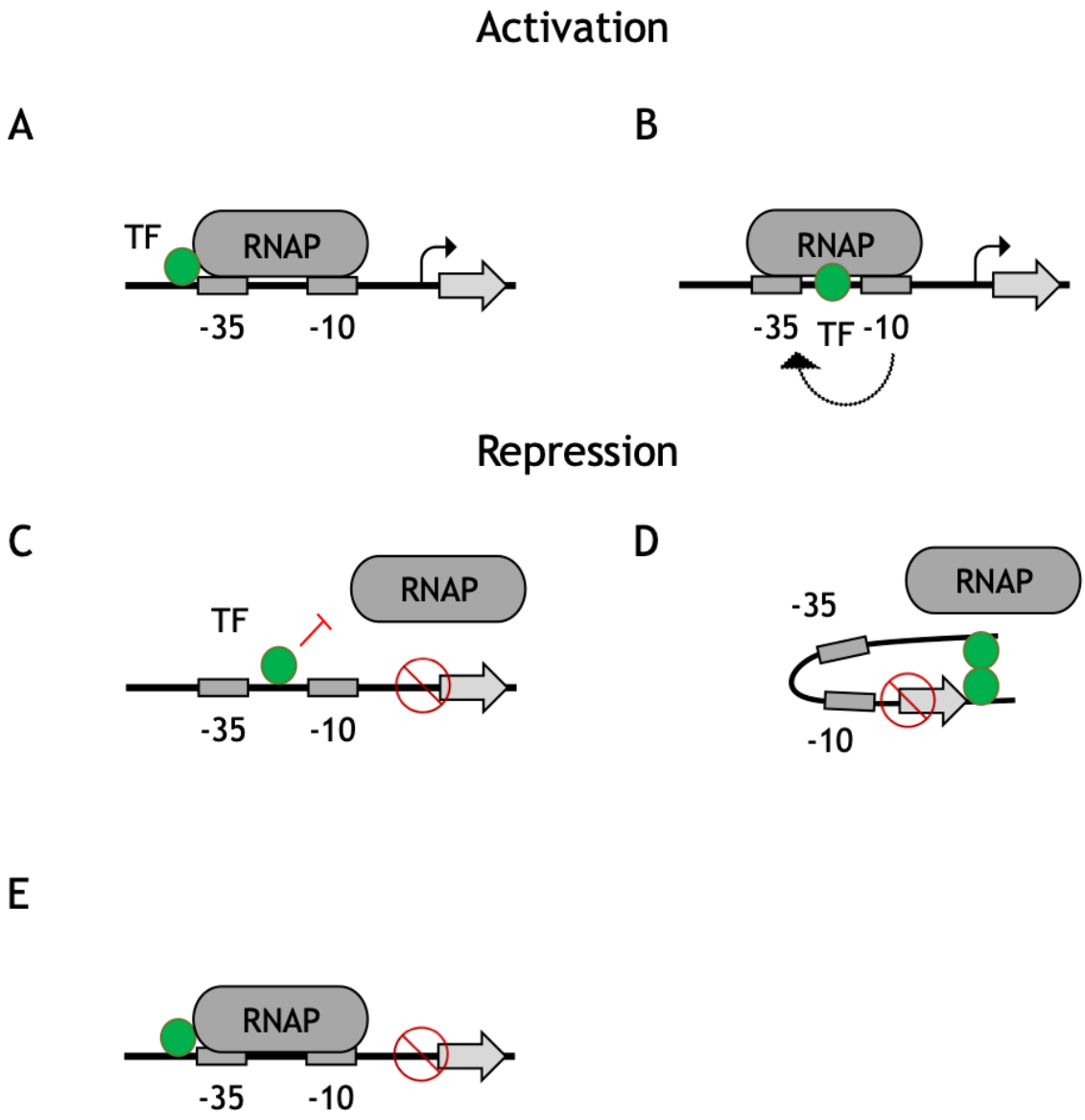


Figure 1-5 The five general mechanisms of gene activation and repression, mediated by TFs. Adapted from Browning, Butala & Busby, 2019. (A) and (B) are

two general methods of TF activation. (A) TF binds to the target DNA and recruits RNAP to the promoter element, enabling trancription. (B) TF binds to the promoter elemnt and adjusts the positioning of the -10 and -35, thus enabling the RNAP to bind and transcription to proceed. (C), (D) and (E) are three general methods of TF repression. (C) TF binds within the promoter element, thereby sterically hindering RNAP from binding, thus repressing transcription. (D) Multiple TFs bind near the promoter element, causing DNA looping, thus occluding RNAP binding and thereby repressing transcription. (E) RNAP binds to the promoter element, however the TF locks it into position, thereby repressing transcription.

Unlike TFs, most NAPs bind to DNA in a promiscuous fashion, often in A:T-rich regions (Grainger *et al.*, 2006). NAPs are in high abundance within the cell, effecting numerous transcriptional responses (Azam *et al.*, 1999). *E. coli* has at least 12 different NAPs, with H-NS being the most well characterised (Babu & Teichmann, 2003a; Martínez-Antonio & Collado-Vides, 2003; reviewed in Dillon & Dorman, 2010). H-NS is often referred to as the 'genome guardian', recognising and binding to A:T-rich regions of DNA, and thus silencing transcription (Grainger *et al.*, 2006; reviewed in Dorman, 2007). As well as affecting gene regulation, NAPs also play a role in the organisation and folding of the nucleoid. *In vitro* studies have revealed that NAPs are able to bend, bridge, wrap and cluster the DNA following binding, with these mechanisms also affecting transcription (Wolf *et al.*, 1999; Dame, Wyman & Goosen, 2000; reviewed in Browning, Grainger & Busby, 2010). Intriguingly however, a recent paper that has redefined the fundamental concepts of transcription initiation suggests that, due to the overlap in functions with TFs, NAPs should no longer be considered a separate class to TFs (Mejía-Almonte *et al.*, 2020).

Two-component systems (TCS) are further regulatory factors of gene expression. Prototypical TCS comprise of a sensor kinase and a response regulator (reviewed in Mitrophanov & Groisman, 2008). The sensor kinase is usually exposed to the environment and responds to environmental signals by modifying the phosphorylated state of the response regulator. The phosphorylation of the response regulator causes a modification in the biochemical properties of the protein and thus enables it to bind to DNA and elicit transcriptional control over genes (reviewed in Mitrophanov & Groisman, 2008).

Regulation through DNA-binding proteins enables a fine-tuning of gene expression in response to environmental signals. This allows bacteria to utilise their genetic material in an appropriate manner, usually enabling them to gain a competitive edge over other bacteria. Although originally TFs were thought to be exclusively proteins, new research has revealed that there are other molecules that play a key role in modulating bacterial gene transcription, including ppGpp and regulatory RNAs (Magnusson *et al.*, 2005; Mejía-Almonte *et al.*, 2020). This further highlights the importance of transcriptional regulation within bacteria.

1.3.4 LysR-type transcriptional regulators (LTTR)

Although TFs are comprised of many distinct protein families, one of the largest groups of prokaryotic TFs are the LysR-type transcriptional regulators (LTTRs) (Pareja *et al.*, 2006). Originally, LTTRs were described as negatively autoregulating themselves and transcriptionally activating a single divergently transcribed gene (Lindquist *et al.*, 1989). However, further research has revealed that LTTRs can be global regulators, affecting the regulation of multiple operons including metabolism, virulence and cell division (reviewed in Maddocks & Oyston, 2008).

LTTRs are typically similar in terms of shape, with the N-terminus comprising of a highly conserved helix-turn-helix (HTH) motif that interacts with the major groove of DNA, and a less-conserved C-terminus that is comprised of a co-inducer binding domain (Henikoff *et al.*, 1988; Muraoka *et al.*, 2003; Schell, 1993). Due to poor solubility, the crystal structures of LTTRs have been notably difficult to solve (Ezezika *et al.*, 2007); however, in 2003, the first full length structure of a LTTR, CbnR, was solved (Muraoka *et al.*, 2003). CnbR was resolved as a tetramer (Muraoka *et al.*, 2003), which corresponded with previous data that LTTRs are often functionally active as dimers and tetramers (McFall *et al.*, 1998).

In transcriptional activation, most LTTRs bind as dimers at two regions within the promoter region, the regulatory binding site (RBS) and the activation binding site (ABS) (McFall *et al.*, 1998). The dimers then interact, forming tetramers, causing the DNA to bend. RNAP is then recruited to the region, although no

transcriptional activity occurs in the absence of the co-inducer. When the co-inducer binds this causes a relaxation of the DNA, allowing activation of transcription (Monroe *et al.*, 1990; McFall *et al.*, 1997; reviewed in Maddocks & Oyston, 2008) (Fig 1-6). There are exceptions to this, however, with the LTTRs Nac and NodD3 not requiring a co-inducer (reviewed in Schell, 1993).

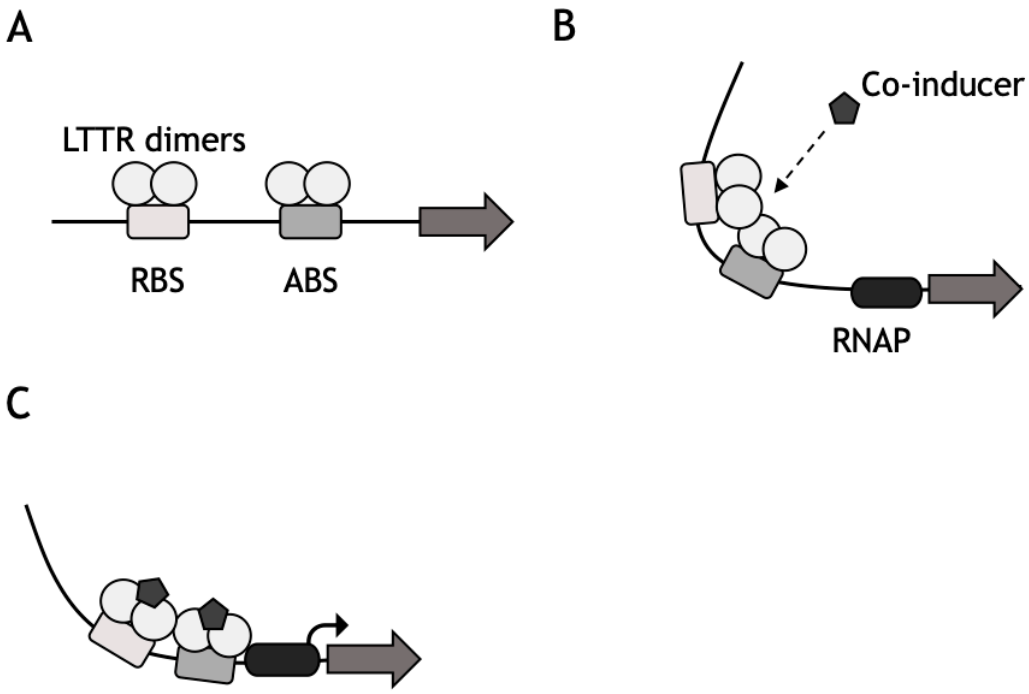


Figure 1-6 A representation of LTTR activation. Adapted from Maddocks & Oyston, 2008. (A) Shows two LTTR dimers binding at the regulatory binding site (RBS) and the activation binding site (ABS). (B) The LTTR dimers interact, forming a tetramer and causing the DNA to bend. RNAP binds at the promoter region, but in the absence of a co-inducer, no transcriptional regulation occurs. (C) The co-inducer binds to the LTTR tetramer, and the DNA bend relaxes, enabling RNAP to interact with the LTTR, thus activating transcription of the gene.

1.3.5 Transcriptional elongation and termination

Although transcriptional regulation occurs most frequently within the initiation phase of transcription, it can also occur in transcriptional elongation and termination. Once the RNAP has 'escaped' the initiation transcription phase, the

transcription cycle proceeds with elongation. The speed and termination of the elongation phase can be influenced by regulatory factors, with NusG and NusA suppressing or enhancing pausing (reviewed in Washburn & Gottesman, 2015).

There are two main types of transcriptional termination; intrinsic termination (therefore template encoded), or termination that requires accessory factors such as Rho and Mfd (reviewed in Washburn & Gottesman, 2015). Intrinsic termination occurs at inverted GC-rich sequence repeats that form an RNA hairpin, which is then followed by a poly-A residue tail (Wilson & Von Hippel, 1995). There are three main proposed models as to how this mechanism works (reviewed in Roberts, 2019), with the 'forward translocation' model suggesting that RNA hairpin formation drives RNAP and the transcription bubble downstream without further RNA transcript elongation, thereby freeing the transcript from RNAP and decreasing the stability of the elongation complex (Santangelo & Roberts, 2004; Yarnell & Roberts, 1999). The intrinsic termination mechanism relies on template encoded RNA hairpins, whereas Rho-dependent termination relies on the ATP-dependent Rho protein. Rho binds to *rut* sites in Rho-dependent terminators and translocates along the RNA 5'-3' to the RNAP (reviewed in Washburn & Gottesman, 2015). This causes the RNAP complex to disassociate from the DNA, terminating transcription; although the mechanism for this is currently unclear.

Regulation of termination can occur in both a positive manner as antiterminators, or a negative manner as anti-antiterminators (reviewed in Henkin, 2000). BglG, a TF of the *E. coli bgl* operon, acts as an antiterminator, binding to the *bgl* leader RNA transcript, blocking the formation of a terminator sequence, thereby allowing expression of the *bgl* operon (Arnster-Choder & Wright, 1993). Conversely, the *trp* RNA-binding attenuation protein (TRAP) in *B. subtilis* acts as an anti-antiterminator of the tryptophan biosynthesis pathway, binding to the nascent *trp* leader transcript, thereby blocking formation of an antiterminator structure, resulting in transcription termination (Babitzke, 1997).

1.3.6 Elucidating the roles of transcription factors

As discussed above, transcriptional regulation can occur within any step of the transcription process. However, it most frequently occurs within the

transcription initiation stage. Encoding DNA-binding proteins that can modulate gene expression in response to environmental signals enables bacteria to survive under fluctuating conditions. Bacterial genomes encode hundreds of TFs, and identifying all of their targets has been a major challenge (reviewed in Wade, 2015). However, with the advent of next generation sequencing, two genome-scale approaches, chromatin immunoprecipitation coupled with next-generation sequencing (ChIP-Seq) and RNA-sequencing (RNA-Seq), have been utilised to map TF regulons (reviewed in Wade, 2015).

ChIP-Seq is a method that is now routinely used in elucidating the genome wide binding of TFs (Barski *et al.*, 2007; Johnson *et al.*, 2007; Mikkelsen *et al.*, 2007; Robertson *et al.*, 2007). TFs are cross-linked to the genome, lysed, and the DNA is fragmented by sonication. The TF-bound DNA is then immunoprecipitated out of the solution and converted into a library for next generation sequencing. The reads generated from this analysis are then mapped back to a reference genome, and the enrichment of bound genomic regions can then be determined against a control sample (reviewed in Wade, 2015). Although this reveals a wealth of information about the TF, it is insufficient to reconstruct the regulon.

RNA-Seq is therefore used to compare RNA levels between the wild-type (WT) and a TF-mutant strain, thus determining the direct and indirect regulon of a TF. For RNA-Seq, RNA is extracted from the samples and undergoes mRNA enrichment or rRNA depletion, followed by cDNA synthesis (reviewed in Stark, Grzelak & Hadfield, 2019). The samples are then converted into a library for next generation sequencing. The reads generated from this analysis are then aligned to the reference genome and are used to analyse differential gene expression (DGE) between samples. Thus, combining the ChIP-Seq and the RNA-Seq data reveals the direct regulon of a TF.

1.4 The role of host metabolites in bacterial virulence

1.4.1 Host metabolites affect bacterial gene expression

The mammalian host is a complex environment, where competition for space and resources is often at its greatest. Rapid responses to environmental cues

enable bacteria to adapt to preferred niches, resulting in maximised competitiveness, usually through the deployment of virulence machinery or by altering nutrient uptake (reviewed in O'Boyle *et al.*, 2020). A prime example of this is the complex environment of the gastrointestinal tract (reviewed in Cameron & Sperandio, 2015). Indeed, within the intestine there are abundant host metabolites, including free fatty acids, amino acids, and vitamins (reviewed in Li *et al.*, 2018).

Pathogenic bacteria, for example EHEC, utilise these host metabolites to modify gene expression, in order to compete against the resident microbiota (reviewed in Cameron & Sperandio, 2015). EHEC deploy a T3SS, which enables them to form attaching and effacing lesions on host epithelia (McDaniel *et al.*, 1995; Moon *et al.*, 1983). The T3SS is encoded on the LEE, which in turn is controlled by a master regulator Ler (McDaniel *et al.*, 1995; Mellies *et al.*, 1999). Expression of *ler* can be directly regulated by the transcriptional regulators Cra and KdpE, which sense the levels of carbon sources available; in gluconeogenic conditions, expression of *ler* is promoted, whereas glycolytic conditions inhibits expression of *ler* and thus the formation of the T3SS is inhibited (Njoroge *et al.*, 2012). Moreover, other host metabolites have been shown to affect the regulation of the LEE in EHEC strains. Indeed, the host metabolite fucose has been demonstrated to effect expression of the LEE (Pacheco *et al.*, 2012). FusKR is a TCS that senses the mucosal sugar fucose, found in abundance in the intestine. In response to sensing fucose, FusK phosphorylates and FusR represses transcription of *ler*, thereby inhibiting formation of the T3SS, speculatively to prevent early expression of the energetically costly production of the T3SS (Pacheco *et al.*, 2012). Further, there are numerous more host signals that can affect the expression of the LEE PAI and subsequently the production of the T3SS (reviewed in Cameron & Sperandio, 2015; Connolly, Finlay & Roe, 2015). The two examples presented here represent how EHEC can choose to deploy a virulence determinant, the T3SS, in response to sensing host metabolite signals, and thus enable deployment and attachment at the opportune moment for maximal effectiveness.

Host-derived free fatty acids can also affect bacterial gene regulation (reviewed in Li *et al.*, 2018). In *Staphylococcus aureus*, the type seven secretion system is activated by *cis*-unsaturated fatty acids (Lopez *et al.*, 2017). Moreover, it is not

solely pathogenic bacteria that regulate gene expression in response to environmental signals. *Campylobacter jejuni* is both a commensal and pathogenic organism that resides within the intestinal tract (Luethy *et al.*, 2017). In order to coordinate expression of genes required for commensal colonisation, *C. jejuni* has been shown to respond to lactate (Luethy *et al.*, 2017). Indeed, in response to lactate, which is present in high concentrations in the avian upper intestinal tract and where *C. jejuni* have been demonstrated to have a reduced ability to colonise, the expression of genes required for commensal colonisation were repressed (Luethy *et al.*, 2017). These examples therefore indicate that adapting gene expression in response to environmental signals is widespread within bacteria, enabling them to survive in the competitive niches of the mammalian and avian host.

1.4.2 Elucidating the regulon of bacterial transcription factors in response to environmental signals

As discussed in 1.3.3, altering gene expression in response to environmental signals is often mediated through TFs or TCS (Fig. 1-7). However, through the advent of next generation sequencing, studying the effect of a TF in response to a signal is no longer restricted to a single region or pathogenicity island (reviewed in Wade, 2015). Indeed, through the combined use of ChIP-Seq and RNA-Seq, a TFs regulon, in response to an environmental signal, can be now elucidated in an unbiased global manner (reviewed in Wade, 2015).

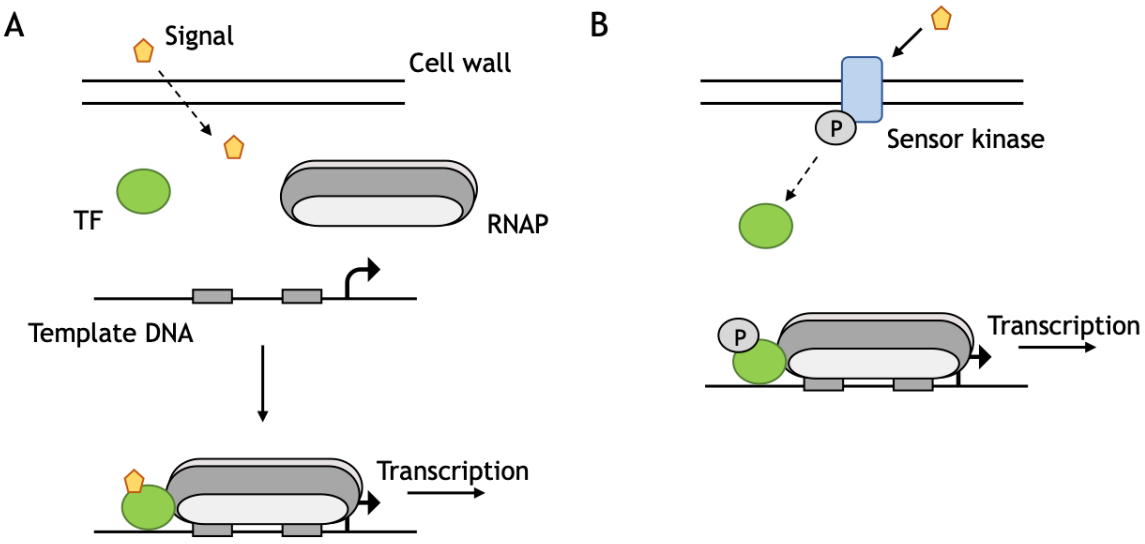


Figure 1-7 TF gene regulation mechanisms in response to environmental signals. Adapted from O'Boyle *et al.*, 2020. (A) Signals are sensed and

recognised by TFs inside the bacterial cell. In this figure, this enables recruitment of the RNAP to the promoter region, activating transcription of target genes. However, TFs activated by environmental signals can also repress transcription of target genes. (B) A signal is sensed by an exposed sensor kinase, that phosphorylates in response. The phosphoryl group is then transferred to a TF, or cognate response regulator, which mediates the expression of target genes.

Recent work on the human pathogen *Bordetella pertussis* has shown how the TCS, BvgAS, integrates virulence and metabolism through the phosphorylation of BvgA (BvgA~P) (Moon *et al.*, 2017). Using RNA-Seq, Moon *et al.*, 2017 elucidated the regulon of BvgA under non-modulating conditions, where BvgA was phosphorylated (BvgA(+)), and modulating conditions, where phosphorylated BvgA was undetectable (Bvg(-)). In this study, BvgA was observed affecting the transcript levels of 550 genes in *B. pertussis* (Moon *et al.*, 2017). It was observed that BvgA~P increased expression of virulence genes including *ptxABDEC*, the pertussis toxin genes and *fimABCD,* the fimbriae genes (Moon *et al.*, 2017). Conversely, when BvgA was unphosphorylated (Bvg(-)), this lead to an upregulation of various metabolic pathways including fatty acid metabolism and the glyoxylate pathway (Moon *et al.*, 2017). These marked differences in regulation suggest that BvgA has adapted to suit the specific lifestyle of the bacterium. Indeed, when *B. pertussis* is in the host, BvgA~P directly activates virulence genes required for colonisation, whereas in the BvgA(-) mode, which has been suggested to be involved in transmission of the pathogen, the metabolic changes could potentially enhance survivability of the pathogen (Moon *et al.*, 2017).

In *Brucella abortus*, VjbR is a LuxR-type transcriptional regulator that has been reported to play a role in expression of known virulence genes (Kleinman *et al.*, 2017). To understand the global role of VjbR in *B. abortus*, ChIP-Seq and RNA-Seq were performed under acidic conditions that mimic the *in vivo* environment *B. abortus* initially encounters within the host. The ChIP-Seq analysis revealed that VjbR was able to bind to 235 regions of DNA, with 37 of those mapped regions correlating with differentially expressed genes (Kleinman *et al.*, 2017). These included genes involved in the *virB* operon (which is responsible for the

type four secretion system) and also genes involved in the electron transport chain (Kleinman *et al.*, 2017). Conversely, a previous study which looked at the direct transcriptional effects of VjbR-mediated regulation in *Brucella melitensis,* in a rich medium, revealed direct regulation of genes *BAB1_0660* and *BAB1_1335* (Uzureau *et al.*, 2010). Although, there were ChIP-Seq peaks in the aforementioned regions in the acidic environment study, there was no transcriptional effects observed, which indicated that VjbR was able to bind under two different environmental conditions but only elicited a transcriptional response in the 'correct' environment (Kleinman *et al.*, 2017).

However, the differences in results observed could also be due to the fact that the studies utilised two different strains of *Brucella*. Indeed, recent work in the Roe laboratory has revealed that even highly conserved TFs have been tailored to suit strain-specific lifestyles. The LTTR, YhaJ, was demonstrated to directly regulate unique gene sets within two *E. coli* pathotypes (Connolly *et al.*, 2019). Using a combination of ChIP-Seq and RNA-Seq the direct regulon of YhaJ was established in UPEC and EHEC (Connolly *et al.*, 2019). In EHEC, YhaJ bound to 23 chromosomal locations and when compared with the RNA-Seq data, was revealed to directly activate expression of T3SS components (Connolly *et al.*, 2019). Conversely, in UPEC, YhaJ bound to 7 chromosomal locations, and was shown to activate expression of T1F, a key virulence determinant (Connolly *et al.*, 2019). These results therefore indicated that YhaJ had been repurposed within *E. coli* to suit strain-specific lifestyles. Bacteria often mediate gene expression in response to environmental signals and cues. The studies discussed here highlight how using a combination of ChIP-Seq and RNA-Seq, a TFs regulon can be elucidated between different environmental conditions, revealing the intricate dynamics of bacterial regulation.

1.5 D-serine and the *dsdCXA* locus

1.5.1 The role of D-amino acids

Host-derived amino acids are often one of the most prevalent metabolites within the gut (Li *et al.*, 2018). Amino acids, except glycine, can be present in two enantiomer states, D- or L-enantiomers, and these chiral forms can be spontaneously converted, usually by racemases (reviewed in Genchi, 2017). L-

amino acids are often used in protein synthesis, as this chiral form is more energetically preferred, although recent reports have indicated the increasing discoveries of D-amino acids in some proteins (Genchi, 2017; Zagon *et al.*, 1994). Intriguingly, although D-amino acids do not typically become incorporated into protein, they have been found to be present in high concentrations in most species and to fulfil specific biological functions (Genchi, 2017).

D-amino acids have been shown to play a role in murine mucosal defence against pathogens (Sasabe *et al.*, 2016). In mammals, D-amino acid oxidase (DAO) is produced at the host-microbe interface in intestinal epithelial cells, in response to microbial synthesised D-amino acids (Sasabe *et al.*, 2016). DAO is a flavoenzyme that can convert D-amino acids into α-keto acids and the antimicrobial product, H_2O_2. *In vitro* studies revealed DAO-produced H_2O_2 impacted the survival of *Vibrio cholerae* and *Vibrio parahaemolyticus*, thus protecting the small intestine from these pathogens (Sasabe *et al.*, 2016). Moreover, D-amino acids have been shown to stimulate M1 macrophages, whilst also promoting survival of intestinal naïve B cells (Suzuki *et al.*, 2021). Furthermore, D-amino acids have also been shown to play a role in disease progression in humans, in both neurological and renal diseases (reviewed in Friedman, 2010).

D-amino acids are derived in mammals in three main ways: conversion of L-amino acids by racemases; release from microbial sources; and dietary consumption (Genchi, 2017). D-amino acids can be obtained from the diet and are highly prevalent in dairy and fermented products, with 100 g of Emmental cheese resulting in an intake of 70 mg of D-amino acids (Zagon *et al.*, 1994). Moreover, processing of food also induces the formation of D-amino acids, with exposure to heat and or changes in pH increasing the concentration (Zagon *et al.*, 1994). However, approximately one third of human D-amino acid content derives from microbial synthesis (reviewed in Friedman, 2010; Genchi, 2017). Microorganisms synthesise D-amino acids, releasing them into the environment or incorporating them into the bacterial cell wall (Lam *et al.*, 2009). Indeed, D-alanine, D-aspartic acid, and D-glutamic acid have all been found in the bacterial peptidoglycan (Zagon *et al.*, 1994). D-amino acids in the peptidoglycan have been demonstrated to provide resistance to proteases, as well contributing

to the architecture of the peptidoglycan (Lam *et al.*, 2009; reviewed in Cava *et al.*, 2011).

The most abundant D-amino acids in mammals are D-serine and D-aspartate (Genchi, 2017). Although the presence of DAO was found in 1935, it was not until the early 1990's that D-serine was found in mammalian tissues (Nishikawa, 2011). D-serine is produced in higher organisms, often converted from L-serine by the serine racemase, in astrocytes and neurons within the brain; although it can also accumulate from dietary sources and intestinal bacteria (Pollegioni & Sacchi, 2010; Zagon *et al.*, 1994). D-serine has been found to be prevalent in multiple niches of the body, including the brain, spinal cord, retina, kidney, bone, and urine (reviewed in Nishikawa, 2011).

1.5.2 The D-serine metabolism locus, *dsdCXA*

Within the human host, bacteria are often competing for space and resources, which has led to the diversification of carbon source utilisation within bacteria (Hibbing *et al.*, 2010). In 1950, Maas and Davis described how, upon exposure to increased levels of D-serine, *E. coli* could not proliferate due to inhibition of the pantothenic acid pathway (Maas & Davis, 1950). However, it was later discovered that *E. coli* K12 carried *dsdA*, a gene encoding a D-serine deaminase, and thus could metabolise D-serine, converting it to pyruvate and ammonia (McFall, 1964). Further work by McFall and colleagues pioneered the early research on the D-serine metabolism locus in *E. coli*, revealing a secondary gene of the locus, *dsdC* (McFall, 1967). *dsdC* encoded a LTTR that was shown to positively regulate *dsdA*, whilst negatively autoregulating itself (McFall & Heincz, 1983; Nørregaard-Madsen *et al.*, 1995). However, it was not until 1995, after DNA sequencing, that the *dsdCXA* locus was fully annotated, concluding a middle gene of the operon, *dsdX,* which was later revealed to be a D-serine transporter (Anfora & Welch, 2006; Nørregaard-Madsen *et al.*, 1995). Thus, the D-serine metabolism pathway was elucidated. In the absence of D-serine, DsdC negatively auto-regulates its own transcription. Upon exposure to D-serine, DsdC induces the transcription of *dsdXA*, thus enabling greater expression of DsdA, which is then able to break down D-serine into pyruvate and ammonia (Nørregaard-Madsen *et al.*, 1995). Pyruvate is then transported into the Krebs cycle, for use as a carbon source. Furthermore, DsdX was revealed to be the

second D-serine transporter, as D-serine is also able to enter the bacterial cell via CycA (Fig. 1-8) (Anfora & Welch, 2006).

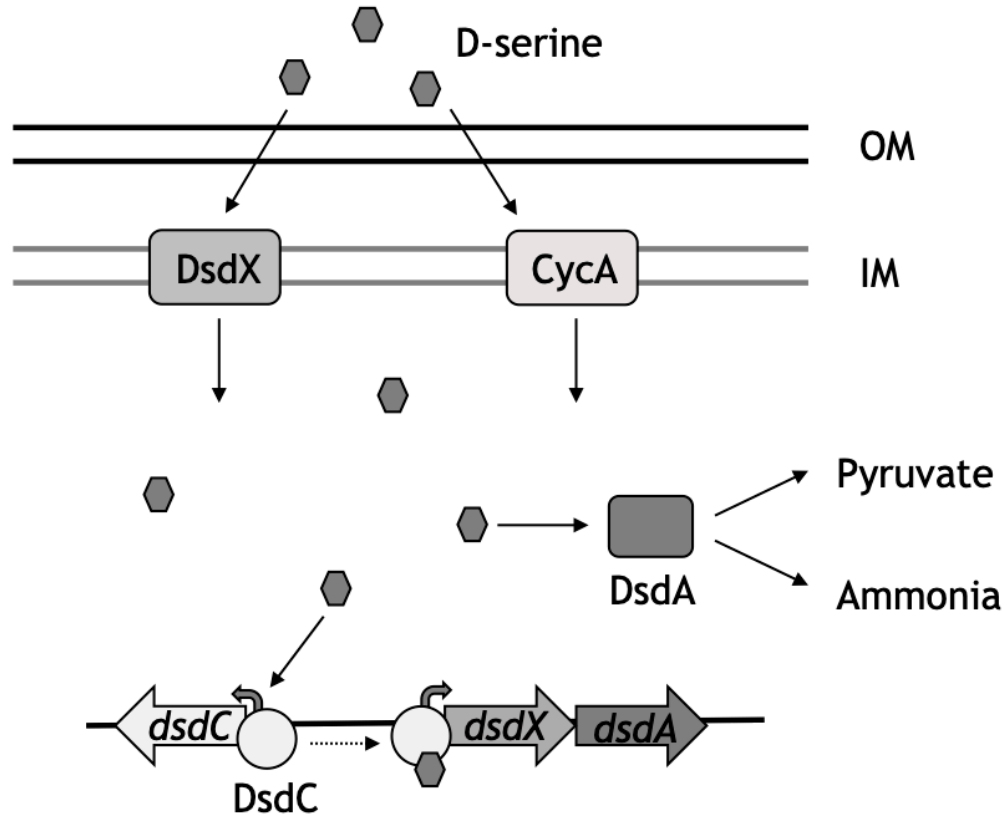


Figure 1-8 Predicted model of the *dsdCXA* locus and subsequent D-serine catabolism. Adapted from Anfora & Welch, 2006. D-serine enters the bacterial cell through DsdX and CycA, the two D-serine transporters. D-serine then interacts with DsdC, causing an increase in expression of *dsdX* and *dsdA*. DsdA breaks down D-serine into ammonia and pyruvate, which subsequently enters the Krebs cycle.

Strikingly, the *dsdCXA* locus was found to be absent in some strains of *E. coli*. A global comparison of three *E. coli* strains revealed that in EHEC, the *dsdCXA* locus had been replaced by genes that encoded for sucrose utilisation, *cscRAKB* (Roesch *et al.*, 2003). D-serine is present in high concentrations within the urinary tract and the brain, however is in low abundance within the intestine (Nishikawa, 2011), indicating that *E. coli* strains have adapted to metabolise the carbon sources available.

1.5.3 D-serine as an environmental signal for bacterial gene regulation

D-amino acids can be utilised in bacteria in three main ways: as a carbon source; as an environmental signal for regulation; and as components of peptides and the bacterial cell wall (reviewed in Cava *et al.*, 2011). Indeed, previous studies have revealed that D-amino acids are involved in numerous regulatory pathways including: regulating the amount of peptidoglycan produced in stationary phase *Vibrio cholerae*; inhibiting spore germination in *Bacillus* species; and inducing the disassembly of *Bacillus subtilis* biofilms (Lam *et al.*, 2009; Kolodkin-Gal *et al.*, 2010; reviewed in Cava *et al.*, 2011).

As discussed above, D-serine can be metabolised in strains that carry the D-serine deaminase, DsdA. Indeed, it was the presence of this regulated and intact *dsdCXA* locus in the UPEC strain CFT073 that drew the focus of Welch and colleagues (Roesch *et al.*, 2003). In an *in vivo* mouse model, it was demonstrated that a Δ*dsdA* mutant "hypercolonised" the bladder and was recovered 300 times more frequently than the WT (Roesch *et al.*, 2003). Furthermore, the Δ*dsdA* mutant was demonstrated to be more motile than the WT (Roesch *et al.*, 2003). Moreover, it was reported in an *in vivo* mouse model, that a Δ*dsdA* strain had upregulated virulence genes that encode P-fimbriae, F1C fimbriae and haemolysin, relative to the WT (Haugen *et al.*, 2007). These results indicated that the accumulation of D-serine affected the gene expression and virulence status of UPEC. However, the authors later established that the Δ*dsdA* mutant phenotypes, and subsequent further study results, were due to an unrecognised secondary mutation in *rpoS*, and *dsdA* neither positively or negatively affected urinary tract colonisation (Hryckowian *et al.*, 2015).

Although UPEC and other *E. coli* strains can metabolise D-serine, many of the diarrheagenic *E. coli* strains have lost this capacity (Roesch *et al.*, 2003). Indeed, Roesch *et al.*, 2003 revealed that out of 60 urinary tract and urosepsis *E. coli* isolates, 49 could metabolise D-serine as the sole carbon source, whereas only 4 out of 74 diarrheal *E. coli* strains were able to. In the EHEC strain EDL933, it was revealed that the *dsdCXA* locus was truncated and had been partially replaced by the genes encoding sucrose utilisation, *cscRAKB* (Roesch *et al.*, 2003). Strikingly, further research revealed that in EHEC, exposure to D-serine

resulted in activation of the SOS response and repression of the T3SS (Connolly *et al.*, 2015). As discussed above, the T3SS enables EHEC strains to form attaching and effacing lesions on host epithelia and is encoded on the LEE (reviewed in Turner, Connolly & Roe, 2018). It was demonstrated that 26 out of the 42 genes that encode the LEE were downregulated in response to D-serine exposure, and furthermore in an *in vitro* cell model there were 77% fewer infected host cells relative to the non-D-serine control (Connolly *et al.*, 2015). As exposure to D-serine had repressed the T3SS irrespective of D-serine catabolism, the authors hypothesised that carriage of both *dsdCXA* and the LEE would be rare (Connolly *et al.*, 2015). Indeed, it was demonstrated that only 1.6% of 1,591 strains carried both the LEE and *dsdCXA*, suggesting a genetic incompatibility between the two loci (Connolly *et al.*, 2015).

Further roles of D-serine as an environmental signal effecting gene regulation have been elucidated, including in the uropathogen *Staphylococcus saprophyticus* (Korte-Berwanger *et al.*, 2013). Exposure to D-serine was demonstrated to upregulate Ssp, a surface-associated lipase and known virulence factor of *S. saprophyticus* (Korte-Berwanger *et al.*, 2013). Furthermore, the Δ*dsdA* mutant was revealed to have a significant disadvantage in comparison to the WT in *in vivo* coinfection experiments (Korte-Berwanger *et al.*, 2013). Moreover, D-serine has been shown to repress gene expression of several virulence associated genes in methicillin resistant *S. aureus* (MRSA) (Iwata *et al.*, 2021). *agrA, icaA, sarS, dltD*, and *sdrD* were all repressed upon exposure to D-serine and thus inhibited the attachment and biofilm formation of MRSA (Iwata *et al.*, 2021). Further, utilisation of D-serine also contributed to the fitness of *Proteus mirabilis* during polymicrobial catheter-associated urinary tract infections (Brauer *et al.*, 2019).

These studies therefore highlight the diverse responses that bacteria undergo in response to exposure to the host metabolite, D-serine. Indeed, recent research has indicated that the transcriptional response to D-serine is distinct even within the same species of bacteria that can utilise D-serine (Connolly *et al.*, 2021). In two pathotypes of *E. coli* that carry the *dsdCXA* locus, UPEC and NMEC, exposure to D-serine induced differential expression of 55 genes in NMEC and 140 genes in UPEC, with only 12 genes commonly differentially regulated in both strains

(Connolly *et al.*, 2021). This indicated that even within strains of the same species that can metabolise D-serine, exposure to the host metabolite results in distinct transcriptional responses.

1.5.4 DsdC, the LysR-type transcriptional regulator

The D-serine metabolism locus comprises of three genes: *dsdC*, the LTTR; *dsdX*, the D-serine inner membrane transporter; and *dsdA*, the D-serine deaminase (Nørregaard-Madsen *et al.*, 1995). In 1967, McFall first described DsdC as a potential regulatory protein that regulated *dsdA*, with further research confirming that DsdC affected the transcription of *dsdA* and not the translation (Heincz & McFall, 1978; McFall, 1967). DsdC was demonstrated to be resolutely required for the activation of *dsdA*, with rate of enzyme synthesis increasing 3,000-5,000 fold upon exposure to D-serine (Bornstein-Forst *et al.*, 1987). Although DsdC has been shown to be the transcriptional regulator of the *dsdCXA* operon in *E. coli*, it is unknown whether DsdC plays any further roles in regulation.

1.6 Aims of project

D-serine is a host metabolite for strains of *E. coli* that can live in extraintestinal niches of the body. D-serine is present in high concentrations within the urinary tract and the brain; however, it is in low abundance within the intestine. Indeed, *E. coli* strains living in the intestine have lost the ability to catabolise D-serine, replacing the D-serine tolerance locus (*dsdCXA*) with a sucrose utilisation locus (*cscRAKB*). NMEC and UPEC are both *E. coli* pathotypes that can metabolise D-serine, carrying the *dsdCXA* locus. Strikingly, NMEC pathotypes have been discovered carrying two copies of the *dsdCXA* locus, potentially indicating its importance for the strain. In addition to being carbon sources, host metabolites can also be environmental signals that can affect gene expression. DsdC is known to regulate the *dsdCXA* locus, however any further regulatory roles are currently unknown. As metabolism and virulence are often interlinked processes, it has been hypothesised that DsdC may regulate other genes in UPEC and NMEC. Thus, the three specific aims of this work were:

1. To elucidate the direct regulon of the TF DsdC, in the presence and absence of D-serine, in two pathotypes of *E. coli*, UPEC (strain CFT073) and NMEC (strain CE10).

2. To elucidate whether there was any significance in CE10 carrying two copies of *dsdC*: *dsdC1* and *dsdC2*.

3. To establish whether DsdC regulation mediated any phenotypes that affected the physiology of CFT073 and CE10.

Chapter 2 describes the materials and methods used in this thesis. Chapter 3 discusses the use of ChIP-Seq and RNA-Seq, which was performed to elucidate the direct regulon of DsdC, in the presence and absence of D-serine. Chapter 4 discusses the validation of the techniques used in Chapter 3 and the annotation of the binding motif of DsdC. Finally, Chapters 5 and 6 delve into the potential phenotypes mediated by DsdC regulation in CFT073 and CE10.

Chapter 2 Materials and methods

2.1 Bacterial strains, plasmids, growth media and chemicals

2.1.1 Strains and plasmids used in these studies

Strain	Characteristics	Source
CFT073	Wild type UPEC O6:H1:K2	Prof Rodney Welch (Welch *et al.*, 2002)
CFT073$^{DsdC\text{-}FLAG}$	UPEC strain with FLAG-tag attached to the *dsdC* gene	Roe lab inventory
CFT073 Δ*dsdC*	UPEC *dsdC* knockout	Roe lab inventory
CE10	Wild type NMEC O7:K1	Prof Kwang Sik Kim (Lu *et al.*, 2011)
CE10$^{DsdC1\text{-}FLAG}$	NMEC strain with FLAG-tag attached to the *dsdC1* gene	Roe lab inventory
CE10 Δ*dsdC1*	NMEC *dsdC1* knockout	Roe lab inventory
CE10$^{DsdC2\text{-}FLAG}$	NMEC strain with FLAG-tag attached to the *dsdC2* gene	Roe lab inventory
CE10 Δ*dsdC2*	NMEC *dsdC2* knockout	Roe lab inventory
CE10 ΔΔ*dsdC1/2*	NMEC *dsdC1*/*dsdC2* knockout	Roe lab inventory
CE10 Δ*neuO*	NMEC *neuO* knockout	This study – Dr James PR Connolly
Nissle 1917	Probiotic *E. coli* strain	Dr Donal Wall (Reister *et al.*, 2014)
Nissle 1917 Δ*dsdC*	Nissle *dsdC* knockout	This study
MG1665	Lab strain K-12 *E. coli*	Roe lab inventory
DH5α	Commercial *E. coli* storage strain	Invitrogen
BL21 DE3	Commercial *E. coli* overexpressing strain	Invitrogen

Table 2-1 Bacterial strains and associated characteristics used in these studies.

Plasmid	Characteristics	Source
pMAL-c5X	Plasmid for overexpression of N-terminal maltose binding protein (MBP) tagged proteins (Amp^R)	New England Biolabs
pMAL-c5X_*dsdC*	MBP-tag overexpression plasmid with CFT073 derived *dsdC* inserted between BamHI and HindIII (Amp^R)	This study
pMAL-c5X_*dsdC1*	MBP-tag overexpression plasmid with CE10 derived *dsdC1* inserted between BamHI and HindIII (Amp^R)	This study
pMAL-c5X_*dsdC2*	MBP-tag overexpression plasmid with CE10 derived *dsdC2* inserted between BamHI and HindIII (Amp^R)	This study
pSR	Plasmid used for DNase foot-printing (Amp^R)	Dr Douglas Browning
pSR_*dsdCX1*	pSR with CE10 derived *dsdCX1* intergenic region inserted between EcoRI and HindIII (Amp^R)	This study
p*dsdA*	pACYC184-*dsdA* complementation construct (Cm^R)	Roesch *et al*., 2003.
pET-28A	Plasmid for overexpression of C-terminal His tagged proteins (Kan^R)	Roe lab inventory
pET-28A_*dsdC1*	His-tag overexpression plasmid with CE10 derived *dsdC1* inserted between BamHI and HindIII (Kan^R)	This study
pKD3	Template plasmid for Lambda Red mutagenesis (Kan^R)	Datsenko and Wanner., 2000.
pKD4	Template plasmid for Lambda Red mutagenesis (Cm^R)	Datsenko and Wanner., 2000.
pKD46	Lambda Red recombinase expressing plasmid (Amp^R)	Datsenko and Wanner., 2000.
pACYC184	Multicopy plasmid (Cm^R)	Roe lab inventory
p*dsdC*	pACYC184 with CFT073 derived *dsdC*	Roe lab inventory

	complementation construct (Cm^R)	
p*dsdC1*	pACYC184 with CE10 derived *dsdC1* complementation construct (Cm^R)	Roe lab inventory
p*dsdC2*	pACYC184 with CE10 derived *dsdC2* complementation construct (Cm^R)	Roe lab inventory
p*neuO*	pACYC184 with CE10 derived *neuO* complementation construct (Cm^R)	This study
p*gapAp_neuO*	pACYC184 with the *gapA* promoter region fused to *neuO* (Cm^R)	This study

Table 2-2 Bacterial plasmids used in this thesis.

2.1.2 Bacterial growth

Bacteria were inoculated using a single colony and grown in 5 ml of Luria-Bertani broth (LB) overnight (16 hours) at 37°C, 200 RPM (New Brunswick Scientific controlled environment incubator shaker). The cultures were then diluted into fresh media the following day at a concentration of 1/100, until the desired OD_{600} had been reached. If growth curves were being undertaken, the bacteria were grown until hour 8.

2.1.3 Storage of strains

For each bacterial strain, a single colony was taken from a LB plate, and grown overnight (16 hours) at 37°C, 200 RPM (New Brunswick Scientific controlled environment incubator shaker). 0.75 ml of this culture was added to a sterile tube with the addition of 1 ml sterile glycerol (40%) and peptone (2%). The stocks were then frozen at -80°C.

2.1.4 Chemicals and reagents

All chemicals used in this thesis were purchased from Merck, Invitrogen, Sigma-Aldrich and ThermoFisher Scientific unless otherwise stated. PCR primers, 1kb Plus DNA ladder and SeeBlue Plus 2 protein standard were all purchased from Invitrogen. Chemiluminescent Western blotting substrates: Femto, Pierce ECL, and Pierce ECL Plus were all purchased from ThermoFisher Scientific. Restriction enzymes were purchased from New England Biolabs. Antibodies: Anti-FLAG, Anti-MBP and Anti-Mouse HRP conjugate were purchased from Sigma-Aldrich; Anti-Rabbit HRP conjugate was purchased from Invitrogen; and Phospho-Histone H2A.X (Ser139/Tyr142) and Beta-tubulin antibodies were purchased from Cell Signalling Technologies. Antibiotics were used at their recommended working concentrations; ampicillin at 100 μg/ml, chloramphenicol at 25 μg/ml and kanamycin at 50 μg/ml.

2.1.5 Growth media and buffers

All growth media and buffers were prepared using either nuclease free water (nfH_2O) or distilled deionised water (ddH_2O) and sterilised either by autoclaving

or filtration (0.2 μM). Solid media was prepared by adding 15 g/L agar to liquid media before autoclaving. All growth medias and buffers were made to 1 litre unless stated otherwise. The relevant components were added to 950 ml ddH_2O, the pH was adjusted, and media topped to 1 litre before sterilisation.

Ingredient	Quantity
Tryptone	10 g
Yeast extract	5 g
NaCl	10 g

Table 2-3 LB recipe; pH 7.5

Ingredient	Quantity
Tryptone	20 g
Yeast extract	5 g
NaCl	0.5 g
KCl (1 M)	2.5 ml
MgSO4 (1 M)	2 ml

Table 2-4 Super optimal broth with catabolite repression (SOC) recipe; pH 7.0. Preceding usage, 5 ml of $MgCl_2$ (2 M) and 20 ml of glucose (1 M) was added.

Ingredient	Quantity
M9 salts (5X)	200 ml

Table 2-5 M9 minimal media recipe; pH 7.5. Preceding usage; 20 ml of 20% Glucose, 2 ml of 1M $MgSO_4$ and 100 μl of 1M $CaCl_2$.

Ingredient	Concentration
Hepes-KOH pH 7.5	50 mM
Sodium deoxycholate	0.1%
Sodium dodecyl sulphate (SDS)	0.1%
Ethylenediaminetetraacetic acid (EDTA)	1 mM
Triton X-100	1%
NaCl	150 mM

Table 2-6 FA lysis buffer (150mM).

Ingredient	Concentration
Hepes-KOH pH 7.5	50 mM
Sodium deoxycholate	0.1%
SDS	0.1%
EDTA	1 mM
Triton X-100	1%
NaCl	500 mM

Table 2-7 FA lysis buffer (500mM).

Ingredient	Concentration
Tris-HCl (pH 8.0)	10 mM
LiCl	250 mM
Sodium deoxycholate	0.5%
Nonidet P-40	0.5%
EDTA	1 mM

Table 2-8 ChIP wash buffer.

Ingredient	Concentration
Tris-HCl (pH 7.5)	50 mM
SDS	1%
EDTA	10 mM

Table 2-9 ChIP elution buffer.

Ingredient	Concentration
Tris-HCl (pH 7.5)	10 mM
EDTA	1 mM

Table 2-10 Tris-EDTA (TE) buffer.

Ingredient	Concentration
Tris-HCl (pH 7.5)	20 mM
NaCl	150 mM

Table 2-11 Tris buffered saline (TBS).

Ingredient	Quantity
10X phosphate buffered saline (PBS)	100 ml
Tween-20	0.4 ml

Table 2-12 Phosphate buffered saline Tween (PBST).

Ingredient	Quantity
Tris	242 g
Acetic acid (glacial)	57.1 ml
EDTA 0.5M (pH 8.0)	100 ml

Table 2-13 Tris-acetate-EDTA (TAE).

Ingredient	Quantity
Glycine	3.75 g
SDS	0.25 g
Tween-20	2.5 ml

Table 2-14 Western blot stripping buffer (pH 2.2).

Ingredient	Quantity
Methanol	500 ml
Coomassie	2 g
Acetic acid	100 ml

Table 2-15 Coomassie blue.

Ingredient	Quantity
Tris (pH 6.8)	1 M
SDS	8.52 g
Bromophenol blue	0.43 g
Glycerol	42.6 ml
β- mercapthenol	4.97 ml

Table 2-16 4X sample buffer (*100 ml volume).

Ingredient	Concentration
Tris	20 mM
NaCl	200 mM
EDTA	1 mM

Table 2-17 MBP binding buffer.

Ingredient	Concentration
Tris	20 mM
NaCl	200 mM
EDTA	1 mM
Maltose	10 mM

Table 2-18 MBP elution buffer.

Ingredient	Concentration
Tris	50 mM
NaCl	200 mM
Imidazole	40 mM
Glycerol	10 %

Table 2-19 His-tag purification buffer.

Ingredient	Concentration
Tris	50 mM
NaCl	200 mM
Imidazole	500 mM
Glycerol	10 %

Table 2-20 His-tag elution buffer.

Ingredient	Quantity
Nutrient Broth Nº2	18 g
Agar	6.3 g

Table 2-21 PB agar.

Ingredient	Concentration
$MgSO_4$	1 mM
$CaCl_2$	4 mM
NaCl	1 mM
Tris-HCl (pH 7.8)	0.05 M

Table 2-22 Phage buffer.

Ingredient	Quantity
Nutrient Broth Nº2	20 g
Agar	3.5 g

Table 2-23 Phage top agar (PTA).

2.2 Molecular and genetic techniques

2.2.1 Polymerase chain reaction (PCR), primers and PCR purification

Primers were designed using Primer3web or NCBI PrimerBLAST. They were designed (unless stated differently) to be around 20 bp long, with a Tm between 55-65ºC. Primers involved in cloning were designed to have the relevant restriction enzyme flanks. Primers were ordered from Life Technologies and came as lyophilised samples and were made to a concentration of 100 μM. Working stocks of 10 μM were prepared for use in PCR reactions.

Primer name	Description	Sequence (5' – 3')
DsdC_FLAG_CHECK_ALL_F	Forward check for DsdC-FLAG tag	GTTTGGCGATATGACGGTGA
DsdC_FLAG_CHECK_UPEC_R	Reverse check for CFT073 DsdC-FLAG tag	TTCGACGCCAAGAAGAAGTG
DsdC1_FLAG_CHECK_NMEC_R	Reverse check for CE10 DsdC1-FLAG tag	TTCGACGCCAAGAAGATGTG
DsdC2_FLAG_CHECK_NMEC_R	Reverse check for CE10 DsdC2-FLAG tag	CGACTCCGGTACATACGACA
DsdC_ChIP_qPCR_1_ALL_F	Forward for ChIP-PCR for the intergenic region between *dsdCX*	CGCAGGCTGACAAACGATAA
DsdC_ChIP_qPCR_1_ALL_R	Reverse for ChIP-PCR for the intergenic region between *dsdCX*	CGGCAAAATTTCAGCTGTAAGG
CitF_ChIP_qPCR_UPEC_F	Forward for ChIP-PCR for CFT073, for a transcriptionally silent gene (*citF*)	CGATGATGGAAAGCGCAGAA
CitF_ChIP_qPCR_NMEC_F	Forward for ChIP-PCR for CE10, for a transcriptionally silent gene (*citF*)	GATGATGGAAAGCGCAGAGG
CitF_ChIP_qPCR_ALL_R	Reverse for ChIP-PCR for a transcriptionally silent gene (*citF*)	ATCACATCGAAATCAGCGCC
MBP_CHECK_DSDC_F	Forward check for pMAL-C5X insertion	CGCGCAGACTAATTCGA
MBP_CHECK_DSDC_R	Reverse check for pMAL-C5X insertion	CCTACTCAGGAGAGCGTTC
DsdC_Clone_U1_MBP_F	Forward for cloning CFT073 *dsdC* with BamHI restriction enzyme	CAGGGATCCGTGATTATGGAACCC
DsdC_Clone_U1/N1_MBP_R	Reverse for cloning CFT073/CE10 *dsdC*/*dsdC1* with HindIII restriction enzyme	CGCAAGCTTGCTCTCCAATATTCGACGCC
DsdC_Clone_N1/N2_MBP_F	Forward for cloning CE10 *dsdC1/2* gene with BamHI restriction enzyme	CAGGGATCCATGGAACCCCTTCG
DsdC_Clone_N2_MBP_R	Reverse for cloning CE10 *dsdC2* with HindIII restriction enzyme	CGCAAGCTTCGACTCCGGTACATACGACA
EMSA_DsdC_CE10_F	Forward for the ChIP peak region for EMSA for *dsdC* region on CE10	CGTGCTCACAACCCAGATTT
EMSA_DsdC_CE10_R	Reverse for the ChIP peak region for EMSA for *dsdC* region on CE10	CCTGGCTGCCACTTCAAAAG
EMSA_EnvR_CE10_F	Forward for the ChIP peak region for EMSA for *envR* region on CE10	CCAAAAGCGCGTTATTTACCC
EMSA_EnvR_CE10_R	Reverse for the ChIP peak region for EMSA for *envR* region on CE10	ACAGAGTGAGAAAATAGCGAAGG
EMSA_FimE_CE10_F	Forward for the ChIP peak region for EMSA for *fimE* region on CE10	GATAGTTGAGATACCAGGGATGG
EMSA_FimE_CE10_R	Reverse for the ChIP peak region for EMSA for *fimE* region on CE10	TCGGTCAAAATTAGCCCAATAGA
BLA_EMSA_F	Forward for the beta-lactamase (*bla*)	CCGCCTCCATCCAGTCTATT
BLA_EMSA_R	Reverse for the beta-lactamase (*bla*)	GGTCGCCGCATACACTATTC
CTX_EMSA_F	Forward for the cholera toxin (*ctx*)	CGGGCAGATTCTAGACCTCC
CTX_EMSA_R	Reverse for the cholera toxin (*ctx*)	TGGATGAGGACTGTATGCCC
UPEC_EMSA_FimB_F	Forward for the ChIP peak region for EMSA for *fimB* region on CFT073	ACCCGAAGTGATGTGGCTAT
UPEC_EMSA_FimB_R	Reverse for the ChIP peak region for EMSA for *fimB* region on CFT073	GATAGTTGAGATACCAGGGATGG
EMSA_WaaV_F	Forward for the ChIP peak region for EMSA for *waaV* region on CE10	TATTCCTGGCAGACGAAACG
EMSA_WaaV_R	Reverse for the ChIP peak region for EMSA for *waaV* region on CE10	AGCCACGAATGAAAAGCAGT
All_DsdC_qPCR_F	Forward to check relative expression of *dsdC* using RT-qPCR	GTTTGGCGATATGACGGTGA
NMEC_UPEC_DsdC1_qPCR_R	Reverse to check relative expression of CFT073/CE10 *dsdC*/*dsdC1* using RT-qPCR	ATAACATTCATATTGACTTACCTG
NMEC_DsdC2_qPCR_R	Reverse to check relative expression of CE10 *dsdC2* using RT-qPCR	ACTGACAAGACTGACATACCTG
gapA_qPCR_F	Forward for housekeeping control (*gapA*) used in RT-qPCR	TTTCCGTGCTGCTCAGAAAC
gapA_qPCR_R	Reverse for housekeeping control (*gapA*) used in RT-qPCR	GGCCGTGAGTGGAGTCATAT
pSR_check_F	Forward check pSR insertion	CCGAAAAGTGCCACCTGAC
pSR_check_R	Reverse check pSR insertion	CGACAAGTTGCTGCGATTCT
dsdC1_insert_pSR_footprint_F	Forward to clone the *dsdCX1* intergenic region into pSR plasmid for footprinting with EcoRI RE	CGCGAATTCCGTGCTCACAACCCAGATTT
dsdC1_insert_pSR_footprint_R	Reverse to clone the *dsdCX1* intergenic region into pSR plasmid for footprinting with HindIII RE	GCGAAGCTTCCTGGCTGCCACTTCAAAAG
waaV_qPCR_F	Forward for *waaV* gene expression using RT-qPCR	TATACGCAAGGCGCAAGGAC
waaV_qPCR_R	Reverse for *waaV* gene expression using RT-qPCR	TCCCGTCTCATGTTGTCTGG
neuB_qPCR_F	Forward for *neuB* gene expression using RT-qPCR	AAAAGCCAAAGAGGCCGGT
neuB_qPCR_R	Reverse for *neuB* gene expression using RT-qPCR	CTGCCTTAGGTGCAATAGCTGA
neuO_qPCR_F	Forward for *neuO* gene expression using RT-qPCR	TCGGTAGAAGAACGACAATAGGT
neuO_qPCR_R	Reverse for *neuO* gene expression using RT-qPCR	ATGCAGTCATGGCCAATTGT
neuO_EMSA_ChIPsite_F	Forward for EMSA to check ChIP peak binding site for *neuO*	CAAGACTCAAGACTCGTTTTCCA
neuO_EMSA_ChIPsite_R	Reverse for EMSA to check ChIP peak binding site for *neuO*	ACGACTTCAAAACCTGCACC

neuO_EMSA_intergenicsite_F	Forward for EMSA to check for intergenic binding site for *neuO*	GCAGCGCAGAGAAATGGATA
neuO_EMSA_intergenicsite_R	Reverse for EMSA to check for intergenic binding site for *neuO*	TCAATGGAAAACGAGTCTTGAGT
neuO_Red_F	Forward for *neuO* lambda red mutagenesis in CE10	TTTTAATGACAAGAAAAGTAACTAGTGGTAAAATAACGTAGGATACTAACGTGTAGGCTGGAGCTGCTTC
neuO_Red_R	Reverse for *neuO* lambda red mutagenesis in CE10	TTTATTACCTTATATTAATATTGGTAATATGTCTGCATGATGTTTTATATCATATGAATATCCTCCTTAG
neuO_184_F	Check forward for *neuO* lambda red mutagenesis in CE10	TGAAGTCAGCCCCATACGATGCCGCTGACTGATAATGACG
neuO_184_R	Check reverse for *neuO* lambda red mutagenesis in CE10	CAATCCATGCCAACCCGTTCCATTACCGCGGTTAATGCCC
EMSA_ClbR_F	Forward for EMSA for validation of ChIP at *clbR* region	TGTTGACTTGGTGTTAGCGG
EMSA_ClbR_R	Reverse for EMSA for validation of ChIP at *clbR* region	GGTGGCTGTATCAATTCATACCC
EMSA_ClbA_F	Forward for EMSA for validation of ChIP at *clbA* region	AGCTTGGTGTCGATATTGAACA
EMSA_ClbA_R	Reverse for EMSA for validation of ChIP at *clbA* region	GGTGATGAGTGGAGAGGCTA
qPCR_ClbA_F	Forward for *clbA* gene expression using RT-qPCR	TAGCCTCTCCACTCATCACC
qPCR_ClbA_R	Reverse for *clbA* gene expression using RT-qPCR	TCAATTCTGCCCATTTGACGA
qPCR_ClbR_F	Forward for *clbR* gene expression using RT-qPCR	ACCCGTTATCTCTGCGTGAA
qPCR_ClbR_R	Reverse for *clbR* gene expression using RT-qPCR	TCTCATTCCTGTTAGCAATGTGT
qPCR_ClbM_F	Forward for *clbM* gene expression using RT-qPCR	GCAGGTATACGACAGGGACA
qPCR_ClbM_R	Reverse for *clbM* gene expression using RT-qPCR	TTACTCGTGTTGTTGCCGTG
DsdC_Red_nissle_F	Forward for *dsdC* lambda red mutagenesis in Nissle 1917	CATCTAAAAATGAAGGTGAATTGAGATATGGTTCACTTTAGCTCACCTTAGTGTAGGCTGGAGCTGCTTC
DsdC_Red_Nissle_R	Reverse for *dsdC* lambda red mutagenesis in Nissle 1917	TGACAAAACAATTCCCATATAAAAATTGCAATTTATAAAGCCAACATACACATATGAATATCCTCCTTAG
DsdC_Nissle_check_F	Check forward for *dsdC* lambda red mutagenesis in Nissle 1917	CGCAGGCTGACAAACGATAA
DsdC_Nissle_check_R	Check reverse for *dsdC* lambda red mutagenesis in Nissle 1917	GCTCTCCAATATTCGACGCC
NeuO_ChIP_qPCR_R	Reverse for ChIP-PCR of the *neuO* region - EMSA_neuO is forward to match	GCCATCGACACTCAACATCA
WaaV_ChIP_qPCR_F	Forward for ChIP-PCR of the *waaV* region	TCAAGTGAAATACCAACATGCAA
WaaV_ChIP_qPCR_R	Reverse for ChIP-PCR of the *waaV* region	GCTCATCGTGTTCTCTTGCC
NeuB_ChIP_qPCR_R	Reverse for ChIP-PCR of the *neuB* region - EMSA_neuB is forward to match	TCCTTATTCTCGATGTCTGCAA
kpsF_EMSA_F	Forward for EMSA for validation of ChIP at *kpsF* region	CGGCCAGATTTAATTCCGCA
kpsF_EMSA_R	Reverse for EMSA for validation of ChIP at *kpsF* region	GCCCACCTATTTAACACACTCC
kpsM_EMSA_F	Forward for EMSA for validation of ChIP at *kpsM* region	ACTGAGGGATGGTGTTGGTT
kpsM_EMSA_R	Reverse for EMSA for validation of ChIP at *kpsM* region	TGCGTAACAACACCTGCAAT
kpsF_qPCR_F	Forward for *kpsF* gene expression using RT-qPCR	GGGCACGTTATTCTTTCGGG
kpsF_qPCR_R	Reverse for *kpsF* gene expression using RT-qPCR	GTAATCATGCCCAGATCGCC
kpsM_qPCR_F	Forward for *kpsM* gene expression using RT-qPCR	GGGGTATTTGTGGGCGATTC
kpsM_qPCR_R	Reverse for *kpsM* gene expression using RT-qPCR	CGGTGCATAACGTAACCCAA
pACYC-184_gibson_F	Forward for pACYC-184 linearisation	TGAATGGAAGCCGGCGGC
pACYC-184_gibson_R	Reverse for pACYC-184 linearisation	CATACACGGTGCCTGACTGC
pACYC_neuO_gibson_F	Forward for *neuO* for gibson assembly	GCAGTCAGGCACCGTGTATGTTATTGCGTGAGCTTCGC
neuO_gapAp_gibson_R	Reverse for *neuO* for gibson assembly	GGTGGAATATATGTCAAGACTCAAGACTCAAG
gapAp_neuO_gibson_F	Forward for *gapA* promoter for gibson assembly	GTCTTGACATATATTCCACCAGCTATTTG
gapAp_pACYC_R	Reverse for *gapA* promoter for gibson assembly	GTGCCGCCGGCTTCCATTCACGTAATTGCCCTTTAAAATTC

Table 2-24 Oligonucleotides used in this study.

The DNA polymerase used was GoTaq Green Master Mix (Promega) or Q5 High-Fidelity 2X Master Mix (NEB). A single colony of bacteria was mixed with 50 μl of nfH_2O and used as a template for the PCR reaction.

Component	Volume
GoTaq Green Master Mix 2X	10 μl
Template DNA	1 μl
Forward and reverse primer (10 μM)	0.5 μl each
nfH_2O	8 μl

Table 2-25 Volumes used for a PCR reaction.

Extension time was altered dependent on the size of PCR product with the general rule of 1 minute per 1 Kb of amplification. Thermocyclers were set for 30 cycles (Eppendorf Mastercycler Nexus gradient).

Step	Temperature (ºC)	Time (min)	Cycles
Initial denaturation	95	5.00	1
Denaturation	95	0.25	30
Annealing	55	0.25	
Extension	72	0.50	
Final extension	72	10.00	1

Table 2-26 Thermocycler conditions for PCR.

DNA was purified using 25-100 μl of PCR product and the QIAQuick PCR purification kit (QIAGEN) to the manufacturer specifications. Purified DNA was eluted in 30 μl of nfH_2O and concentrations measured using the NanoDrop DS-11+ Spectrophotometer (DeNovix).

2.2.2 Plasmid purification

5 ml of overnight culture containing the plasmid were centrifuged at 13,000 g for 5 minutes. Using the QIAprep Spin Miniprep kit (QIAGEN), plasmid DNA purification was performed to the manufacturer specifications. Plasmid DNA was eluted in 30 μl of nfH_2O and concentrations measured using the NanoDrop DS-11+ Spectrophotometer (DeNovix).

Plasmids were also purified using the QIAprep Spin Maxiprep kit (QIAGEN) and were performed to the manufacturer specifications. Plasmid DNA was eluted in 30 μl of nfH_2O.

2.2.3 Restriction enzyme digest and DNA ligation

For restriction enzyme digests, 1 μg of purified DNA (2.2.1) was incubated with 1 μl of the respective restriction enzyme (NEB), 2 μl of respective buffer and made up to an overall volume of 20 μl using nfH_2O. These were incubated at 37°C for 1 hour, unless specified differently. The digests were run on a 1% agarose TAE gel and the samples purified using the QIAquick Gel extraction kit (QIAGEN) (2.2.6).

Digested DNA was ligated into linearised plasmid at a ratio of 3:1 respectively. 1 μl of T4 ligase (NEB) and 2 μl of T4 buffer (NEB) were added to the reaction with nfH_2O to a total volume of 10 μl. These were incubated at room temperature for 2 hours. 5 μl of the sample was used for transformation and added to 50 μl of competent cells and transformed either using heat shock transformation (2.2.5) or electrocompetent cell transformation (2.2.6). These were plated onto antibiotic plates for selected screening and positive colonies were confirmed by both PCR and DNA sequencing (Eurofins).

2.2.4 Preparation of competent cells

A single colony was inoculated in 5 ml of LB and grown to an OD_{600} of 0.4 at 37°C, 200 RPM (New Brunswick Scientific controlled environment incubator shaker). The cells were centrifuged at 3,750 g for 5 minutes at 4°C. The supernatant was removed, and the cells were washed and centrifuged at 15,000 RPM three times in 1 ml of ice cold ddH_2O. Cells were then suspended in 50 μl of ice cold ddH_2O and used in transformations.

2.2.5 Heat shock transformation

50 μl of competent cells was mixed with 5 μl of ligated plasmid and incubated on ice for 30 minutes. The samples were then heat shocked at 42°C for 30 seconds and samples placed back on ice for 2 minutes. 950 μl of pre-warmed

SOC was added to each reaction and incubated for 1 hour at 37°C. 100 μl of the reaction was then plated out onto the antibiotic selective plate.

2.2.6 Electrocompetent cell transformation

50 μl of competent cells was mixed with 5 μl of ligated plasmid and incubated on ice for 5 minutes. The samples were added to a pre-chilled electroporation cuvette and shocked at 2,500 volts in an Eporator electroporator (Eppendorf). 950 μl of pre-warmed SOC was added to each reaction and incubated for 1 hour at 37°C. 100 μl of the reaction was then plated out onto the antibiotic selective plate.

2.2.7 Agarose gel electrophoresis and gel purification of DNA

1% agarose TAE gels were used for gel electrophoresis. 10,000X GelRed Nucleic acid gel stain (Cambridge Bioscience) was added to the liquid agarose. A 1kb Plus DNA ladder (Invitrogen) was used as the molecular weight marker. Gels were run for 40 minutes at 100 volts. The DNA was visualised using a UV transilluminator.

DNA that needed to be purified from the agarose gels was extracted using QIAquick Gel extraction kit (QIAGEN) to the manufacturer specifications, and purified DNA was eluted in 30 μl of nfH_2O.

2.2.8 Phenol:chloroform extraction and ethanol precipitation

For phenol:chloroform extractions, samples were made up to 400 μl and mixed with an equal volume of phenol:chloroform:isoamillic alcohol 25:24:1 (PCIA) (Sigma Aldrich). These were centrifuged at maximum speed for 5 minutes. After the centrifugation, the top layer was dispensed into a new Eppendorf and 400 μl of chloroform:isolamillic alcohol 24:1 (CIA) (Sigma Aldrich) was added. The samples were then inverted several times carefully and centrifuged at maximum speed for 1 minute. After centrifugation, the top layer was again dispensed into a new Eppendorf and 40 μl of sodium acetate, 1 μl of GlycoBlue Coprecipitant 15

mg/ml (ThermoFisher Scientific) and 800 μl of 100% ethanol was added to the samples. The samples were then inverted several times and the stored in the -80ºC for 1 hour. The samples were then centrifuged at maximum speed for 20 minutes at 4ºC. The supernatant was discarded and 1 ml of 70% ethanol was added. The samples were centrifuged for a further 5 minutes at maximum speed. The supernatant was then carefully removed, and the samples were left to airdry. The pellets were then suspended in appropriate media, nfH_2O or TE buffer.

2.2.9 Lambda Red genetic recombination

This method was carried out as described in Datsenko and Wanner, 2000. For this technique, concentrated PCR product of the resistance cassette, from either pKD3 or pKD4, was generated with 50 bp overhangs directly adjacent to the gene of interest to be knocked out. Four 50 μl PCR reactions were set up and checked by gel electrophoresis, before they were pooled. A phenol:chloroform extraction and ethanol precipitation (2.2.8) was then completed on the PCR product.

Strains desired for gene knock outs were transformed with pKD46 prior to the recombination, with recovery at 30ºC. A transformed colony was inoculated into 5 ml of LB broth, 100 μg/ml ampicillin and 100 μl of 1M arabinose, the inducer of the Lambda Red system. Cultures were grown at 30ºC at 200 RPM (New Brunswick Scientific Innova 44 incubator shaker series) until an OD_{600} of 0.3 was reached. Cultures were then centrifuged at 3,750 g for 5 minutes and the supernatant removed. The pellet was suspended three times in 1 ml of ice-cold ddH_2O and centrifuged at 10,000 g for 1 minute, with the supernatant removed. The pellets were suspended in a final volume of 50 μl of ddH_2O and mixed with 1 μg of PCR purified DNA. These samples were then added to pre-chilled electroporation cuvettes and electroporated. 500 μl of SOC was immediately added and the samples recovered at 37ºC at 200 RPM (New Brunswick Scientific Innova 44 incubator shaker series). 200 μl of the reaction was plated out onto selection plates and incubated at 37ºC. Colonies were then screened by PCR and gel electrophoresis. Successful recombinants were then plated on LB at 42ºC to get rid of the resistance cassette, leaving a clean mutant.

2.2.10 Gibson assembly

Primers were designed for Gibson assembly using the NEBuilder assembly tool (NEB). Sequences were amplified by PCR with the relevant overhangs and pACYC184 was linearised. Gibson assemblies were carried out using a Gibson Assembly Cloning Kit (NEB) to the manufacturer specifications. The following reaction was set up on ice; 0.5 pmols of vector:insert at a ratio of 1:1, 10 μl of NEBuilder HiFi DNA Assembly Master Mix, and nfH_2O to a total volume of 20 μl. Samples were incubated in a thermocycler at 50°C for 15 minutes. 2 μl of the sample was transformed into DH5α using the heat shock method (2.2.5). The samples were then sequenced to ensure for correct alignment and insertion.

2.3 Genomic and transcriptomic techniques

2.3.1 Chromatin immunoprecipitation (ChIP)

The ChIP experiment was based on methods previously established by Bonocora and Wade, (2015) with minor alterations. Prior to the ChIP experiment commencing, protein A sepharose beads (CL4B; GE Healthcare), in a 50% TBS slurry, were pre-blocked overnight with 0.1% and 1% bovine serum albumin (BSA).

For the ChIP, duplicate CFT073$^{DsdC\text{-}FLAG}$, CE10$^{DsdC1\text{-}FLAG}$ and CE10$^{DsdC2\text{-}FLAG}$ strains were grown in 50mL of M9 minimal media for 5 hours at 200 RPM (New Brunswick Scientific Innova 44 incubator shaker series). Further, duplicate CFT073$^{DsdC\text{-}FLAG}$, CE10$^{DsdC1\text{-}FLAG}$ and CE10$^{DsdC2\text{-}FLAG}$ strains were grown in 50mL of M9 minimal media for 3 hours at 200 RPM (New Brunswick Scientific Innova 44 incubator shaker series), before they were spiked with 1 mM D-serine, and grown for a further 2 hours. Non-tagged CFT073 WT and CE10 WT samples were also grown in parallel, to be used as negative control in the ChIP (Table 2-27).

ChIP sample	ChIP sample name
CFT073 WT	NT01
CE10 WT	NT02
CFT073$^{DsdC\text{-}FLAG}$ + 1mM D-ser	NT03
CFT073$^{DsdC\text{-}FLAG}$ + 1mM D-ser	NT04
CFT073$^{DsdC\text{-}FLAG}$	NT05
CFT073$^{DsdC\text{-}FLAG}$	NT06
CE10$^{DsdC1\text{-}FLAG}$ + 1mM D-ser	NT07
CE10$^{DsdC1\text{-}FLAG}$ + 1mM D-ser	NT08
CE10$^{DsdC1\text{-}FLAG}$	NT09
CE10$^{DsdC1\text{-}FLAG}$	NT10
CE10$^{DsdC2\text{-}FLAG}$ + 1mM D-ser	NT11
CE10$^{DsdC2\text{-}FLAG}$ + 1mM D-ser	NT12
CE10$^{DsdC2\text{-}FLAG}$	NT13
CE10$^{DsdC2\text{-}FLAG}$	NT14

Table 2-27 ChIP samples and nomenclatures given.

Samples were crosslinked with 1% formaldehyde for 20 minutes before the reaction was stopped by addition of 0.5 M glycine. The samples were centrifuged, washed with TBS, and suspended in FA lysis buffer with 4 mg/ml lysozyme and incubated at 37ºC for 30 minutes. Samples were then sonicated for 2x15 cycle runs (30 seconds on/off) (Bioruptor Pico; Diagenode). The samples were then centrifuged at maximum speed for 5 minutes and the chromatin (supernatant) was removed. A 20μL input sample was saved at this point for ChIP-PCR.

For immunoprecipitation of the samples, 60 μl of 1% BSA Protein A sepharose bead slurry was added to the chromatin and rotated for 2 hours at 4ºC. This step was performed to reduce background noise. The lysate was then removed by centrifugation at 4,000 g for 1 minute. 60 μl of 0.1% BSA Protein A sepharose bead slurry was then added to the lysate, along with 4 μl of M2 anti-FLAG antibody, and rotated overnight at 4ºC. The protein A Sepharose beads are able to bind to the anti-FLAG antibody, which in turn has bound to the DsdC^{FLAG}-DNA.

The beads were then pelleted by centrifugation at 4,000 g for 1 minute, the lysate removed, and the beads suspended in 700 μl of FA lysis buffer (150 mM NaCl) and transferred to a Spin-X column (Corning). These columns were then rotated at room temperature for 3 minutes, before centrifugation for 1 minute

at 4,000 g, and removal of the supernatant. 2 further washes, rotations and centrifugation with FA lysis buffer (150 mM NaCl) were performed. This was followed by one wash, rotation and centrifugation with FA lysis buffer (500 mM NaCl), followed by one wash, rotation and centrifugation with ChIP wash buffer. A final wash, rotation and centrifugation in TE buffer was performed before columns were transferred to dolphin nosed tubes. The supernatant was incubated with 100 μl of ChIP elution buffer at 65ºC for 10 minutes. Samples were then eluted by centrifugation at 4,000 g for 1 min. The samples, and the inputs collected earlier, were then de-crosslinked by boiling for 10 minutes. The supernatants were then concentrated using the phenol:chloroform extraction method, followed by ethanol precipitation described in 2.2.8. The ChIP DNA and inputs were then air-dried before resuspending in 12 μl nfH_2O.

2.3.2 Chromatin immunoprecipitation coupled with next generation sequencing (ChIP-Seq)

The ChIP samples were then sent to the University of Glasgow Polyomics where DNA concentration was measured using a Qubit HS DNA kit (ThermoFisher Scientific). Due to low DNA concentration, another set of duplicate ChIP samples were obtained for CFT073 WT, CFT073$^{DsdC\text{-}FLAG}$ and CE10$^{DsdC2\text{-}FLAG}$, with and without D-serine, and the ChIP samples were then pooled together before sequencing (Table 2-28).

ChIP sample	ChIP sample name
CFT073 WT	NT01, NT01B, NT01C
CE10 WT	NT02
CFT073$^{DsdC\text{-}FLAG}$ + 1mM D-ser	NT03, NT03B
CFT073$^{DsdC\text{-}FLAG}$ + 1mM D-ser	NT04, NT04B
CFT073$^{DsdC\text{-}FLAG}$	NT05, NT05C
CFT073$^{DsdC\text{-}FLAG}$	NT06, NT06B
CE10$^{DsdC1\text{-}FLAG}$ + 1mM D-ser	NT07
CE10$^{DsdC1\text{-}FLAG}$ + 1mM D-ser	NT08
CE10$^{DsdC1\text{-}FLAG}$	NT09
CE10$^{DsdC1\text{-}FLAG}$	NT10
CE10$^{DsdC2\text{-}FLAG}$ + 1mM D-ser	NT11, NT11B
CE10$^{DsdC2\text{-}FLAG}$ + 1mM D-ser	NT12, NT12B
CE10$^{DsdC2\text{-}FLAG}$	NT13, NT13B
CE10$^{DsdC2\text{-}FLAG}$	NT14, NT14B

Table 2-28 Pooled ChIP samples and their nomenclatures.

University of Glasgow Polyomics then generated a ChIP-Seq library which was prepared using a NEBNext Ultra II DNA library prep kit for Illumina (NEB). The libraries were quantified using the Qubit HS DNA kit (ThermoFisher Scientific). The profiles and the size of the libraries were analysed on the Bioanalyser High Sensitivity DNA ChIP (Agilent). Illumina next generation sequencing was performed on the samples using an Illumina NextSeq 500 platform (75 bp length; single end).

2.3.3 ChIP-Seq analysis

Reads were quality assessed (minimum Phred threshold of 20) with FastQC (Babraham Bioinformatics) before importing into CLC Genomics Workbench 7. Raw fastq files were aligned to the CFT073 and CE10 reference genomes (NCBI accession numbers respectively; NC_004431.1 and CP003034). The ChIP-Seq analysis tool software was used, and the $DsdC^{FLAG}$-tagged samples were then aligned against the mock ChIP WT control samples to establish enrichment. The maximum *p*-value for calling enriched peaks was set to ≤ 0.05.

Peaks were called from two biological replicates and checked manually to ensure they were the correct bimodal peak shape. Regions that did not conform were omitted. As ChIP-Seq is known to predict false positives, it was decided that each peak that was called would be manually checked (from the lowest *p*-value generated) until there were at least 10 'peaks' in a row that did not conform to the bimodal peak shape and then the manual search would cease.

2.3.4 ChIP-PCR

Prior to library generation, ChIP-PCR was done to ensure enrichment of target DNA over the non-FLAG tagged control. The conditions used are described in Table 2-29. This enrichment was measured by calculating the signal over the background using fold enrichment $2^{-\Delta\Delta CT}$ ($\Delta CT = CT^{ChIP} - CT^{CONTROL}$). The region between *dsdCX* was used as it was the only verified binding site of DsdC. Regions of *citF* were used as controls for non-specific enrichment.

Cycle step	Temperature (ºC)	Time (minutes)	Cycles
Initial denaturation	95	1	1
Denaturation	95	0.25	40
Extension	60	0.5	

Table 2-29 Thermocycler conditions for ChIP PCR.

2.3.5 RNA extraction and DNase treatment

RNA extraction was carried out using the PureLink RNA Mini Kit (ThermoFisher Scientific). CFT073, CE10, and corresponding isogenic Δ*dsdC* strains, were grown in triplicate, in M9 minimal media for 5 hours at 37ºC, 200 RPM (New Brunswick Scientific Innova 44 incubator shaker series). Further, CFT073, CE10, and corresponding isogenic Δ*dsdC* strains, were grown in triplicate, in M9 minimal media for 3 hours at 37ºC, 200 RPM (New Brunswick Scientific Innova 44 incubator shaker series), before being spiked with 1 mM D-serine and grown for a further 2 hours. The OD_{600} of the samples was adjusted to 1.0. 1 ml of culture was then centrifuged for 10 minutes at maximum speed and the supernatant discarded. The pellet was suspended in 1 ml of RNA protect (QIAGEN) and incubated at room temperature for 10 minutes. The samples were centrifuged for a further 10 minutes at maximum speed and the supernatant discarded. To the samples, 100 μl of TE solution containing lysozyme (10 mg/ml) was added with 0.5 μl 10% SDS and vortexed until the pellet was suspended. These were left to incubate for 5 minutes at room temperature. 350 μl of lysis buffer with 1% 2-mercaptoethanol was added to the samples and vortexed to mix. 250 μl of 100% ethanol was added to the samples and vortexed vigorously to remove any visible precipitation.

The samples were then transferred to a PureLink RNA Mini Kit spin column and centrifuged at maximum speed for 15 seconds. The flow through was discarded and 700 μl of Wash Buffer 1 was added to the same column and centrifuged at maximum speed for 15 seconds. The flow through was discarded again and 500 μl of Wash Buffer 2 was added to the same column and centrifuged at maximum speed for 15 seconds. This was repeated again with 500 μl of Wash Buffer 2. The column was then centrifuged for 1 minute at maximum speed, the supernatant discarded, and the column was inserted into a new recovery tube. 100 μl of

nfH_2O was added to the centre of the column and incubated at room temperature for 1 minute. The column was then centrifuged for 1 minute at maximum speed. The samples were then DNase treated. 4 μl of TURBO DNase (ThermoFisher Scientific) and 10 μl of 10X TURBO DNase buffer was added to each sample. The samples were centrifuged at maximum speed for 30 seconds, mixed gently and incubated for 60 minutes at 37°C. During the 60 minutes incubation, the samples were vortexed and centrifuged every 15 minutes.

The samples were then analysed on a NanoDrop DS-11+ Spectrophotometer (DeNovix) and by electrophoresis on an agarose gel. Readings between 1.8 and 2.0 of 260/280 absorbance were accepted values on the NanoDrop. Further analysis was done by a *groEL* PCR check to ensure that there was no DNA present in the RNA samples.

The RNA from the samples was then concentrated using phenol:chloroform extraction method followed by ethanol precipitation outlined in 2.2.8. The samples were then airdried and suspended in 100 μl of TE buffer. The samples were then normalised to a total concentration of 10 ng/μl in 100 μl.

2.3.6 Quantitative real time PCR (RT-qPCR)

RNA samples were extracted as detailed above (2.3.5) and normalised to a total concentration of 10 ng/μl. The LunaScript RT SuperMix kit (NEB) was used for the cDNA synthesis of the RNA samples. For each reaction: 2 μl of 5X LunaScript RT SuperMix, 7 μl of nfH_2O, and 1 μl of RNA template was used. These samples were put on the thermocycler (Eppendorf Mastercycler Nexus gradient) under the following conditions (Table 2-30).

Cycle step	Temperature (°C)	Time (minutes)
Annealing	25	2
cDNA synthesis	55	10
Heat inactivation	95	1

Table 2-30 Thermocycler conditions for LunaScript RT SuperMix cDNA synthesis.

This was then followed by RT-qPCR performed on a CFX-Connect Real-Time PCR detection system (BIORAD) using the Luna Universal qPCR Master Mix kit (NEB). 1

μl of cDNA was added to 10 μl of 2X Luna Universal qPCR Master Mix, 1 μl of forward and reverse primers (10 μM) respectively, and 7 μl of nfH_2O. The reactions were performed in duplicate and each gene that was analysed was performed in biological triplicate. Two controls were used for the RT-qPCR, a no reverse transcriptase control and a no template control.

Cycle step	Temperature (ºC)	Time (minutes)	Cycles
Initial denaturation	95	1	1
Denaturation	95	0.25	40
Extension	60	0.5	

Table 2-31 Thermocycler conditions for RT-qPCR.

All primers used in RT-qPCR were checked for efficiency prior to the experiment, using the thermocycler conditions described in Table 2-31. Primers are detailed in Table 2-24. 5 standards were made using template cDNA of 100, 20, 4, 0.8 and 0.16 ng/μl. Primers were only used if they had efficiency between 90-110%. The data was then analysed using the CFX-Connect BIORAD software, according to the $2^{-\Delta\Delta CT}$ method.

2.3.7 Transcriptome profiling by RNA-Seq

RNA samples were extracted as detailed above (2.3.5). RNA quality was assessed by Agilent Bioanalyzer 2100. Samples were subjected to ribosomal depletion using MICROBExpress (ThermoFisher Scientific) according to the manufacturer's instructions. Library preparation and sequencing was carried out at the University of Glasgow Polyomics facility. Sequencing libraries were prepared with the TrueSeq Stranded mRNA Library Prep kit (Illumina) according to manufacturer's instructions. Sequencing was carried out on the Illumina NextSeq 500 platform with at least 10 million 75 bp single end reads being obtained. Reads were quality assessed (minimum Phred threshold of 20) with FastQC (Babraham Bioinformatics) before importing into CLC Genomics Workbench (Qiagen) and mapping to the CFT073 and CE10 reference genome and plasmids respectively (NCBI accession number: NC_004431.1, CP003034, CP003035, CP003036, CP003037, CP003038) using default CLC mapping parameters. Differential expression was assessed using the empirical analysis of differential expression (EdgeR) with genes displaying absolute fold changes of ≥

1.5; ≤ -1.5 and having a false-discovery rate corrected *p*-value of ≤0.05 being considered. Pairwise comparison of WT and either Δ*dsdC* or ΔΔ*dsdC1/2* was conducted to identify any changes in gene expression caused by the removal of the transcription factor DsdC. Pairwise comparison of $WT^{D\text{-}ser}$ and either $\Delta dsdC^{D\text{-}ser}$ or $\Delta\Delta dsdC1/2^{D\text{-}ser}$ was used to identify the extent gene expression in the presence of D-serine. Further pairwise comparisons were done with $WT^{D\text{-}ser}$ and either $\Delta dsdC^{D\text{-}ser}$ + pDsdA, or $\Delta\Delta dsdC1/2^{D\text{-}ser}$ + pDsdA1/2 to identify gene expression caused by DsdC and the inducement of D-serine, and not by the toxic effects of D-serine accumulation.

RNA-Sequencing was completed prior to this PhD commencing and the raw data is available on the European Nucleotide Archive (ENA) under accession numbers:

Strain	ENA accession numbers
UPEC	ERS4281315 - ERS4281325, ERS4281353
NMEC	ERS4281326 - ERS4281334, ERS4281354 - ERS4281356

Table 2-32 European Nucleotide Archive (ENA) accession numbers for RNA-Sequencing data. The accession numbers on ENA for the RNA-Seq data "Establishing the global regulatory role of DsdC in UPEC and NMEC".

2.3.8 Sanger sequencing

For samples less than 1 Kb that required sequencing, primers were designed around the sequence of interest. The sequence was then amplified using PCR and purified using a QIAquick PCR purification kit (QIAGEN). Purified PCR products and primers were then sent to Eurofins Genomics for Sanger sequencing. Results were returned as FASTA files and Clustal Omega (EMBL-EBI) was used for multiple sequence alignment.

2.4 Biochemical techniques

2.4.1 SDS-PAGE and Coomassie staining

Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) and Coomassie staining were used to analyse whole cell and purified protein samples, using the NuPAGE system (Invitrogen). The samples were grown under ChIP-Seq growth conditions; 3 hours grown in M9 minimal media at 37°C, 200

RPM (New Brunswick Scientific Innova 44 incubator shaker series), spiked with or without 1 mM of D-serine and left to grow for a further 2 hours. The samples were normalised using OD_{600} and centrifuged at 4ºC for 3 minutes at 10,000 g. The supernatant was discarded, and the pellet was suspended in 75 μl of ddH_2O and 25 μl of sample buffer (Table 2-16). Samples were then heated for 10 minutes at 95^{0}C.

An alternative method was also used. The samples were grown up and normalised using OD_{600}, before being centrifuged at 10,000 g for 3 minutes. The supernatant was discarded and 100 μl of BugBuster Protein Extraction Reagent (Merck) was added to the sample. The samples were then vortexed every 10 minutes for 30 minutes. The samples were centrifuged again for 3 minutes at 10,000 g. The samples were then standardised again using a bicinchoninic acid (BCA) protein assay (Sigma Aldrich). 25 μl of 4X NuPAGE LDS sample buffer (Invitrogen) was added to the samples which were then heated for 10 minutes at 95^{0}C.

Pre-cast NuPAGE 4-12% Bis-Tris gels (Invitrogen) were used. 15 μl of sample was loaded onto the gel and run in NuPAGE MES buffer (Invitrogen) at 180 volts for 40 minutes, until the dye front had reached the end of the gel. SeeBlue Plus 2 was used as the molecular weight ladder. SDS-PAGE gels were then either transferred for Western blotting (2.4.4) or Coomassie stained. The gels were stained for 1 hour in Coomassie blue stain (Table 2-15) and destained overnight in ddH_2O. Coomassie stained SDS-PAGE gels were then imaged on a transilluminator (BIO-RAD ChemiDoc MP Imaging system).

2.4.2 Protein overexpression of recombinant proteins

The CE10 and CFT073 *dsdC* genes were cloned into the maltose binding protein (MBP) overexpression plasmid pMAL-C5X and used for overexpression in *E. coli* BL21 DE3 cells. Several trial over-expression methods were tested before large scale over-expression and purification commenced. Single colonies were used to inoculate LB and were incubated for 16 hours at 37ºC, 200 RPM (New Brunswick Scientific controlled environment incubator shaker). These overnight cultures were then used to inoculate fresh LB and grown to an OD_{600} of 0.6 at 200 RPM, 37ºC (New Brunswick Scientific Innova 44 incubator shaker series). The samples

were then induced with 1 mM isopropyl ß-D-1-thiogalactopyranoside (IPTG) and left to grow for a further 3 hours at 37ºC, 200 RPM. The samples were then centrifuged at 7,000 g for 10 minutes and the supernatant removed. The pellet was suspended in 75 μl of nfH_2O and 25 μl of 4X sample buffer (Table 2-16) and heated to 95ºC for 10 minutes. Samples were then run on pre-cast NuPAGE 4-12% Bis-Tris gels and Coomassie stained (2.4.1) or Western blotted (2.4.5). The CE10 *dsdC1* gene was also cloned into the overexpression plasmid pET-28A and used for overexpression in *E. coli* BL21 DE3 cells.

2.4.3 Protein purification of recombinant proteins

Prior to purification of the MBP-tagged proteins, the plasmids were freshly transformed into BL21 DE3. A single colony was then inoculated and left overnight at 37ºC at 200 RPM (New Brunswick Scientific controlled environment incubator shaker). The overnight culture was then diluted into 1L of LB and antibiotic and grown to an OD_{600} of 0.6 at 37ºC, 200 RPM (New Brunswick Scientific Innova 44 incubator shaker series). 1 mM of IPTG was added and left to grow for a further 3 hours, 225 RPM, at 37ºC. The samples were centrifuged for 20 minutes at 5,000 g and the supernatant removed. The samples were centrifuged again, at 7,500 g for 10 minutes to remove any final supernatant. To the pellet, 35 mg of lysozyme was added with cOmplete EDTA-free protease inhibitor cocktail (Promega), DNase I and 35 ml of MBP binding buffer (Table 2-17). The sample was sonicated on ice, 10 seconds ON and 30 seconds OFF for 20 cycles and centrifuged for 40 minutes at 15,000 g. The supernatant was removed and filtered through a 0.22 μM filter. 15 μl of the sample was removed at this stage to use as a control. The sample was then run through the MBP HP column (GE Healthcare) using a MasterFlex L/S (Cole-Parmer). The flow-through was collected to use as a control. The column was washed with 5 column volumes (CVs) of MBP binding buffer (Table 2-17), followed by 5CVs of MBP elution buffer (Table 2-18), eluting into 2 mL fractions. 5 μl of 4X sample buffer (Table 2-16) was added to 15 μL of each of the fractions and heated to 95ºC for 10 minutes. Samples were then run on pre-cast NuPAGE 4-12% Bis-Tris gels and Coomassie stained (2.4.1) in order to determine which fraction contained the purified recombinant protein. These samples were then aliquoted and kept for long term storage at -80ºC.

Prior to purification of the His-tagged protein, the same protocol was used as above, with the exception of 35 ml of His binding buffer (Table 2-19) was added to the pellet instead of MBP binding buffer. The sample was then either run through the HisTrap HP column (GE Healthcare) using a MasterFlex L/S (Cole-Parmer) or a HisTrap HP column (GE Healthcare) using an AKTA Start (University of Glasgow Protein Purification Service). For the MasterFlex L/S the flow-through was collected to use as a control. The column was washed with 5 column volumes (CVs) of His binding buffer (Table 2-19), followed by 5CVs of His elution buffer (Table 2-20), eluting into 2 mL fractions. 5 μl of 4X sample buffer (Table 2-16) was added to 15 μL of each of the fractions and heated to 95°C for 10 minutes. Samples were then run on pre-cast NuPAGE 4-12% Bis-Tris gels (Invitrogen) (2.4.1) and Coomassie stained (2.4.1) in order to determine which fraction contained the purified recombinant protein. These samples were then aliquoted and kept for long term storage at -80°C. For the AKTA Start, this was performed by the University of Glasgow Protein Purification Service. Fractions were run on pre-cast NuPAGE 4-12% Bis-Tris gels and Coomassie stained (2.4.1).

2.4.4 BCA protein assay

To measure the concentration of the recombinant protein, a BCA Protein Assay was used (Pierce, ThermoFisher Scientific). This was completed using the standard microplate protocol according to the manufacturer specifications. Briefly, standards of BSA were made to concentrations between 25 μg/ml to 2000 μg/ml. A BCA working reagent was made by mixing 50 parts of BCA Reagent A with 1 part of BCA Reagent B. 25 μl of each standard and recombinant protein was added to a 96 well microplate in technical duplicates. 200 μl of working reagent was added to each of the wells and the plate was briefly vortexed. The plate was then covered and incubated for 30 minutes at 37°C. Absorbance was measured on a FLUOstar Optima plate reader (BMG Labtech) at 562 nm. A standard curve was prepared by plotting the 562 nm measurement for each BSA standard against its concentration in μg/ml. The standard curve was then used to measure the relative concentration of the recombinant protein.

2.4.5 Using Factor Xa to cleave the MBP-DsdC recombinant protein

The recombinant protein was cleaved of its MBP tag. Using 1 mg/ml of recombinant protein, 200 μg/ml of Factor Xa Protease (NEB) was added to the sample. The samples were left at room temperature and 5 μl of sample was removed at 2-, 4-, 6- and 24-hours intervals. 1.5 μl of 4X sample buffer was added to the samples, heated for 10 minutes, and run on pre-cast NuPAGE 4-12% Bis-Tris gels and Coomassie stained (2.4.1). A control using no Factor Xa protease was used to check for complete cleavage.

2.4.6 Western blotting

Following SDS-PAGE (2.4.1) the gel was then transferred to either an Amersham Protran NC Nitrocellulose membrane (ThermoFisher Scientific) or Immobilon-P polyvinylidene fluoride (PVDF) membrane (Millipore) using the XCell SureLock Transfer system (Invitrogen). The transfer was run at 30 volts for 60 minutes in NuPAGE transfer buffer (Novex). The membrane was then blocked in 5% milk (Marvel) in PBST for 60 minutes. Primary antibody was added in 1% milk in PBST and left at room temperature rocking for 60 minutes. Membranes were washed 3 times in PBST for 10 minutes before being incubated with the secondary antibody for 60 minutes at room temperature. Membranes were washed again 3 times in PBST. Either Pierce ECL Western Blotting Substrate, SuperSignal West Femto Maximum Sensitivity Substrate or Pierce ECL Plus Western Blotting Substrate was added to the membrane for 5 minutes. These were transferred to either the transilluminator or the C-digit (Li-Cor) and exposed; or exposed to Medical X-ray Blue film (Carestream) and developed using a film processor.

Antibody	Source	Primary antibody concentration	Secondary antibody concentration
FLAG	Mouse	1/1,000	1/2,000
GroEL	Rabbit	1/10,000	1/10,000
DnaK	Mouse	1/3,000	1/6,000
MBP	Mouse	1/2,000	1/5,000
Phospho-Histone H2A.X (Ser139/Tyr142)	Rabbit	1/1,000	1/2,500
Beta tubulin	Rabbit	1/10,000	1/10,000

Table 2-33 Antibody concentrations used for Western blotting.

2.4.7 Electrophoretic mobility shift assay (EMSA)

The EMSA assays were completed using a DIG gel shift kit (2nd generation) to the manufacturer specifications using DIG wash and block buffer set (Roche) and performed in duplicate. Purified MBP-tagged protein (2.4.3), was used for this experiment. Promoter regions, which were analysed from the ChIP-Seq data set, thought to contain DsdC DNA-binding sites were amplified using PCR and confirmed by agarose gel electrophoresis. These were then purified by QIAquick gel extraction kit (QIAGEN) for use as binding probes. The EMSA probes were then DIG-labelled as to the manufacturer specifications, 100 ng of purified DNA was adjusted to a final volume of 10 μL in sterile distilled water. 4 μL labelling buffer (5X), 4 μL $CoCl_2$ solution, 1 μL DIG-ddUTP solution and 1 μL of terminal transferase (400 U) was added to the probes on ice, mixed and centrifuged briefly. The probes were then incubated at 37^{0}C for 15 minutes before being returned to the ice. The reaction was stopped by adding 2 μL of 0.2 M EDTA (pH 8.0) and 3 μL of sterile distilled water. This gave the labelled probes a final concentration of 4 ng/μL. The efficiency of probe labelling was tested by using serial dilutions and spotting 1 μL of DIG labelled probe onto positively charged nylon membrane (Roche). The probe was cross-linked to the membrane using UV light for 20 minutes. The membrane was washed in DIG washing buffer and incubated for 5 minutes at room temperature. The membrane was blocked for 30 minutes using DIG blocking solution and then blocked for a further 30 minutes using DIG antibody solution. Membranes were washed twice for 15 minutes in DIG wash buffer and then incubated in DIG detection buffer for 2 minutes. Membranes were then placed on cling film and 1 ml of CSPD working solution was applied to the membrane. Cling film was immediately placed on top covering the membrane and left to incubate at room temperature for 5 minutes. The excess liquid was then lightly squeezed out of the cling film and sealed. The sealed membranes were then incubated for 10 minutes at 37ºC and exposed on the transilluminator (BIO-RAD Chemidoc).

For the EMSAs: 4 μL of binding buffer, 1 μL of poly [d(I-C)], 1 μL of poly L-lysine, 1 μL of DIG labelled probe and 11 μL of nfH_2O, were added for each reaction. Differing concentrations of purified protein was added to the reactions. These were mixed carefully on ice and incubated at room temperature for 15 minutes.

The samples were placed back onto ice and 5 μL of loading buffer was added to each reaction. 20 μL of sample was then loaded onto a pre-run NOVEX 6% retardation gel (Invitrogen) and ran in 0.5X TBE buffer at 100 V for 1 hour and 45 minutes. The gel was then transferred onto a positively charged nylon membrane (Roche), using the XCell SureLock Transfer system, at 30 V for 1 hour. The membranes were then cross-linked using UV light for 20 minutes. The protocol then follows the same procedure as stated above for the probes.

2.4.8 DNase I footprinting

This was performed at the University of Birmingham under the supervision of Dr Douglas Browning, who then completed the experiment. Briefly, the *dsdCX* intergenic promoter region fragment was cloned into the pSR vector using EcoRI and HindIII. Fragments were sequenced to ensure correct insertion. pSR-*dsdCX* was purified using a QIAGEN MaxiPrep kit (detailed in 2.2.2) with a concentration of at least 100 μg of DNA needed for DNase I footprinting. The fragments were then cut using HindIII and incubated for 3 hours at 37ºC, before being phosphotase treated with calf intestinal phosphatase (CIP). The fragments were then precipitated using a phenol:chloroform extraction followed by an ethanol precipitation (outlined in 2.2.8). Samples were suspended in 50 μl of TE. The fragments were then further cut with AatII, incubated for 3 hours at 37ºC. The samples were then run on a 7.5% acrylamide gel. If the band was the appropriate size, based upon on a molecular ladder, it was cut from the gel and the DNA was electroeluted from the gel slice. Electroelution occurred by preparing dialysis tubing, inserting the gel slice into the tubing and adding 200 μL of 0.1X TBE. These were then run for 20 minutes at 40 mA. The supernatant was removed from the tubing and placed in an Eppendorf. The tube was washed with a further 200 μL of sterile H_2O and the supernatant removed and combined with the sample. The fragments were precipitated using a phenol:chloroform extraction followed by an ethanol precipitation (outlined in 2.2.8). 0.5 μL of sample was run on a 7.5% acrylamide gel to check for DNA. The fragments were then able to be radiolabelled. For each 20 μL reaction: 8 μL of DNA fragment, 8 μL of sterile H_2O, 2 μL of PNK buffer, 1 μL of ATP γ P32, and 1 μL of T4 polynucleotide kinase (NEB) was used. This was incubated at 37ºC for 30 minutes. The unincorporated nucleotides were then removed by passing the

samples through a Sephadex G-50 column. The filtration of the DNA fragments from the unincorporated nucleotides was done by adding 200 μL of Sephadex G-50 beads to a spin column. This was allowed to settle for 5 minutes, before centrifugation at low speed (3.5 g) occurred for 2 minutes. The DNA fragments were then pipetted onto the beads and centrifuged for 2 minutes (3.5 g). The supernatant was kept as the ^{32}P-end radiolabelled DNA fragments, and the columns disposed of appropriately.

For DNase I footprinting, each 20 μL reaction was made up as follows: 0.2 μL of ^{32}P-end labelled DNA fragment, 2 μL of 10X HEPES buffer, 1 μL of 10 mg/ml^{-1} BSA, 1 μL of 0.6 mg/ml^{-1} Herring sperm DNA, varying concentrations of DsdC, and H_2O to bring the volume up to 20 μL. Samples were then incubated for 20 minutes at 37°C and 2 μL of DNase I was added to the sample for 40 seconds before the reaction was stopped with 200 μL of DNase I stop solution. The samples were then precipitated using phenol:chloroform extraction followed by ethanol precipitation (outlined in 2.2.8). The samples were suspended in loading buffer and heated to 90°C for 2 minutes. The samples were loaded onto a 6% denaturing gel with a GA ladder and visualised on a Bio-Rad PMI imager.

2.5 Phenotypical assays

2.5.1 Silver staining

Samples were prepared for silver staining using the ChIP-Seq growth conditions; 3 hours grown in M9 minimal media at 37°C, 200 RPM, spiked with or without 1 mM of D-serine and left to grow for a further 2 hours (New Brunswick Scientific Innova 44 incubator shaker series). Samples were centrifuged for 3 minutes at 12,000 g. 100 μL of BugBuster was added to each of the samples and left to incubate at room temperature for 30 minutes, vortexing every 10 minutes. The samples were centrifuged at 12,000 g for 3 minutes and 25 μL of 4X LDS NuPAGE sample buffer was added. Samples were heated to 95°C for 10 minutes. Proteinase K (NEB) was added to the samples at a concentration of 25 μg and incubated at 60°C for 1 hour. The samples were then run on a pre-cast NuPAGE 4-12% Bis-Tris gel (2.4.1).

Silver staining was carried out using the SilverQuest Silver Staining kit (ThermoFisher Scientific) to the manufacturer specifications. The gel was washed briefly in ddH_2O before 100 ml of fixative solution was added for 20 minutes. The gel was then washed in 30% ethanol for 10 minutes, before 100 ml of sensitising solution was added for a further 10 minutes. 30% ethanol was added to the gel for 10 minutes, before the gel was washed in ddH_2O for 10 minutes. The gel was incubated for 15 minutes in 100 ml of staining solution and washed in ddH_2O for 1 minute. 100 ml of developing solution was added to the gel and incubated for 4-8 minutes until bands started to appear. 10 ml of stopper was immediately to the gel and the gel was incubated for a further 10 minutes being gently agitated. The gel was washed for a final time in ddH_2O and imaged on the transilluminator. All experiments were completed in triplicate.

2.5.2 Serum killing assay

This protocol was adapted from Sarkar *et al.*, 2014. The *E. coli* strains were grown overnight in LB at 37°C, 200 RPM (New Brunswick Scientific controlled environment incubator shaker). The cultures were washed twice in PBS and suspended to an OD_{600} of 0.8. 50 μl of each culture was mixed with 50 μl of human sera (Life Science Group) and incubated at 37°C for 90 minutes at 120 RPM (New Brunswick Scientific controlled environment incubator shaker). This was serially diluted and plated on LB plates. As a control 50 μl of each culture was also mixed with 50 μl of heat inactivated serum, where the serum was incubated at 56°C for 30 minutes prior to the experiment. This was to ensure the complement cascade was inactivated. Samples were also mixed separately with 50 μl of PBS and serially diluted. This experiment was performed in triplicate. CFU/ml was calculated as:

$$\frac{CFU}{ml} = \frac{number\ of\ colonies}{dilution\ of\ plate\ counted\ x\ volume\ of\ lysate\ plated}$$

2.5.3 K1 bacteriophage titre and plaque assay

E. coli strains were grown in LB at 37°C, 200 RPM for 16 hours (New Brunswick Scientific controlled environment incubator shaker). The strains were diluted

1/100 in M9 minimal media and grown until an OD_{600} of 0.6 was reached. 100 μl of bacterial culture was added to 3 ml of warm PTA agar, spread onto PB plates, and allowed to dry. Serial dilutions of K1 bacteriophage (SSI Diagnostica) were done in phage buffer and 10 μl of each dilution was spotted onto the plates in technical replicates. PFU/ml was calculated as:

$$\frac{PFU}{ml} = \frac{number\ of\ plaques}{dilution\ of\ plate\ counted\ x\ volume\ of\ lysate\ plated}$$

Plaque size was also measured using ImageJ software (Fiji). All experiments were completed in triplicate.

2.5.4 K1 bacteriophage killing assay

For the bacteriophage killing assay, *E. coli* strains were grown in LB at 37°C, 200 RPM for 16 hours (New Brunswick Scientific controlled environment incubator shaker). The strains were diluted 1/100 in M9 minimal media in a 96 well microplate and grown for 3 hours in a humidity chamber at 37°C, 200 RPM (New Brunswick Scientific Innova 44 incubator shaker series). 1 μl of K1 phage (SSI Diagnostica) was added to the strains and the OD_{600} was measured every 30 minutes or hour for 3 hours using a FLUOstar Optima plate reader (BMG Labtech). All experiments were completed in triplicate.

2.6 *In vitro* cell infection models and microscopy

2.6.1 HeLa cell culture maintenance

HeLa cell lines were cultured and maintained in Dulbecco's Minimal Eagles Medium (DMEM), containing 20% (v/v) FBS (foetal bovine serum) and 20 mM L-glutamine, at 37°C, 5% CO_2. Cells were cultured in T75 vented flasks and kept until passage 25. The cells were split after the flasks had reached 80% confluency. For splitting, 2 ml of Trypsin-EDTA was added to the flasks and left at 37°C, 5% CO_2 for 5 minutes. 8 ml of pre-warmed DMEM media was added to the flasks and the cells were removed from the flask. Cells were counted using a haemocytometer.

2.6.2 Infection assays on HeLa cell lines

For the genotoxin assays on the HeLa cell lines, 24-well tissue culture plates were coated in a 0.5% (v/v) collagen for 30 minutes prior to seeding. The collagen solution was removed, and the cells were washed once with PBS. 40,000 cells were seeded per well and left for 24 hours, at 37°C, 5% CO_2 to form a confluent monolayer. The Nissle strains were grown in minimal essential medium, HEPES modification (MEM-HEPES) media, with or without 1 mM D-serine for 4.5 hours at 200 RPM, 37°C and normalised to an OD_{600} of 0.1 (New Brunswick Scientific Innova 44 incubator shaker series). For an MOI of 400; 200 μl of bacteria was added to each well, with 300 μl of MEM-HEPES, supplemented with 1% (v/v) L-glutamine. The cells were centrifuged at 200 g for 2 minutes and incubated for 4 hours at 37°C, 5% CO_2. The cells were washed once with PBS and the media replaced with MEM-HEPES, supplemented with 1% (v/v) L-glutamine and 50 μg/ml gentamicin. The cells were incubated for a further 4 hours at 37°C, 5% CO_2. 100 μl of 1X sample buffer (Table 2-16) was added to each well and left for 5 minutes. The cells were removed and boiled for 10 minutes for Western blot analysis (2.4.6). To visualise the genotoxic effect on the host cells via Western blotting, phosphorylated Histone H2A.X (Cell Signalling Technologies) antibodies were used.

Additionally, cells were also used in immunofluorescence microscopy to visualise genotoxic effects. The infection was carried out as above, with some minor changes. The cells were seeded at 40,000 cells per well onto 0.5% (v/v) collagen-coated coverslips and left for 24 hours at 37°C, 5% CO_2 to form a confluent monolayer. The Nissle strains were grown in MEM-HEPES, with or without 1 mM D-serine for 4.5 hours at 200 RPM, 37°C and normalised to an OD_{600} of 0.1 (New Brunswick Scientific Innova 44 incubator shaker series). For an MOI of 400; 200 μl of bacteria was added to each well, with 300 μl of MEM-HEPES, supplemented with 1% (v/v) L-glutamine. The cells were centrifuged at 200 g for 2 minutes and incubated for 4 hours at 37°C, 5% CO_2. The cells were washed once with PBS and the media replaced with MEM-HEPES, supplemented with 1% (v/v) L-glutamine and 50 μg/ml gentamicin. The cells were incubated for a further 48 hours at 37°C, 5% CO_2. The cells were washed twice with PBS and 250 μl of 4% paraformaldehyde (PFA) (w/v) was added to each well and left for 20

minutes. The cells were washed three times with PBS and 250 μl of 0.1% Triton X-100 was added to each well and left for 5 minutes. Host cell actin was stained with AlexaFluor-555 Phalloiden (Invitrogen), which was added to each well and left to rock in darkness for 1 hour. The cells were washed again three times with PBS and the coverslips were mounted onto 4 μl of VECTASHIELD (Vector Laboratories) on microscope slides. These were left to dry for a few minutes and secured using clear nail varnish. Slides were imaged on a Zeiss M1 Axioskopp microscope and deconvolved using Zen Pro software.

2.6.3 hCMEC/D3 cell culture maintenance

hCMEC/D3 (Merck), the blood-brain barrier cell line, were cultured and maintained in hCMEC/D3 ENDOGro-LS Complete medium (Merck), containing ENDOGro Basal Medium, ENDOGro-LS Supplement (0.2% v/v), rh epidermal growth factor (5 ng/ml), ascorbic acid (50 μg/ml), L-glutamine (10 mM), hydrocortisone hemisuccinate (1 μg/ml), heparin sulfate (0.75 U/ml), FBS (2% v/v), and recombinant human fibroblast growth factor (1 ng/ml). A 1:20 Collagen T1 Rat Tail (Sigma) solution was used to coat the T75 flasks prior to seeding. Cells were cultured in T75 vented flasks, at 37°C, 5% CO_2 and kept until passage 10. For splitting, 2 ml of Trypsin-EDTA was added to the flasks and left at 37°C, 5% CO_2 for 5 minutes. 8 ml of pre-warmed hCMEC/D3 ENDOGro-LS Complete media was added to the flasks and the cells were removed. Cells were counted using a haemocytometer.

2.6.4 Adhesion and invasion assays on hCMEC cell lines

For the adhesion and invasion assays on the hCMEC cell lines; 24-well tissue culture plates were coated in a 0.5% (v/v) collagen solution for 1 hour prior to seeding and kept at 37°C. The collagen solution was removed, and wells were washed once with PBS. 40,000 cells were seeded per well and left for 24 hours to form a confluent monolayer. The CE10 strains were grown in M9 minimal media for 5 hours, with or without the addition of 1 mM D-serine spiked in at hour 3, at 37°C, 200 RPM, and normalised to an OD_{600} of 0.1 (New Brunswick Scientific Innova 44 incubator shaker series). For an MOI of 100; 50 μl of bacteria was added to each well with 450 μl of hCMEC/D3 ENDOGro-LS Complete media.

The cells were centrifuged at 200 g for 2 minutes and incubated for 2 hours at 37°C, 5% CO_2.

For the adhesion assay, the cells were washed three times with 500 µl of PBS, with the washes kept to serially dilute in PBS and calculate the CFU/ml. 300 µl of 1% Triton X-100 was added to the wells and left for 5 minutes. The cells were removed from the wells and serially diluted in PBS and the CFU/ml was calculated. Bacterial counts were analysed in GraphPad Prism 8.

For the invasion assay, the cells were washed three times with 500 µl of PBS, with the washes kept to serially dilute in PBS and calculate the CFU/ml. 500 µl of hCMEC/D3 ENDOGro-LS Complete media with 100 µg/ml of gentamicin was added to the cells and left to incubate at 37°C, 5% CO_2 for a further hour. 300 µl of 1% Triton X-100 was added to the wells and left for 5 minutes. The cells were removed from the wells and serially diluted in PBS and the CFU/ml was calculated. Bacterial counts were analysed in GraphPad Prism 8.

2.7 Bioinformatic analysis

2.7.1 Bioinformatic analysis and databases used

DNA and protein sequences were obtained using NCBI. Multiple sequence alignments were done using NCBI BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi) or Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/). Venny 2.1.0 was used for Venn diagram analysis (https://bioinfogp.cnb.csic.es/tools/venny/). Reverse Complement was used to convert a DNA sequence into its reverse complement (https://www.bioinformatics.org/sms/rev_comp.html). Bioline was used for six-frame translation of nucleotide sequences (https://www.bioline.com/media/calculator/01_13.html). SnapGene was used to visualise plasmid maps and restriction cloning. GraphPad Prism 8 was used to create diagrams. Figures were made using Microsoft PowerPoint.

2.7.2 ChIP-Seq and RNA-Seq bioinformatic analysis

ChIP and RNA-Seq analysis was carried out on CLC Genomics Workbench 7 (CLC Bio). Functional characteristics were assigned to predicted targets using UniPROT Knowledgebase and the GO functions (https://www.uniprot.org). MEME was used to predict DNA-protein binding motifs (https://meme-suite.org/meme/).

2.7.3 Statistical analysis

Statistical analysis was carried out on GraphPad Prism unless stated otherwise. For analysis between 2 samples, an unpaired Student *t*-test was used. RT-qPCR analysis was computed using CFX-Connect BIORAD software, according to the $2^{-\Delta\Delta CT}$ method.

Chapter 3 Investigating the global binding of DsdC and its role on gene expression in pathogenic *Escherichia coli*

3.1 Introduction

The capacity to utilise different carbon sources contributes to the ability of ExPEC to thrive in niches of the body that are otherwise less favourable. D-serine is a host metabolite present in the body, particularly in the urinary tract and the brain. Recent works have described the numerous ways in which D-serine can affect the physiology of many bacteria, including inhibiting the formation of the T3SS in EHEC, inhibiting adherence and biofilm production in MRSA and contributing to the increased fitness of *P. mirabilis* in catheter-associated UTIs (Brauer *et al.*, 2019; Connolly *et al.*, 2015; Iwata *et al.*, 2021).

DsdC is a LTTR that regulates the D-serine metabolism locus in *E. coli* by upregulating the *dsdXA* operon whilst negatively autoregulating itself (Nørregaard-Madsen *et al.*, 1995). The *dsdCXA* locus is important for ExPEC, with 344 out of 427 strains from the B2 clade positively carrying the *dsdCXA* locus (Connolly *et al.*, 2015). Moreover, *E. coli* K1 strains often carry two unlinked copies of *dsdCXA*, with 73% of 41 K1 strains carrying two copies; thus highlighting the importance of D-serine metabolism to these strains (Moritz & Welch, 2006). DsdC is therefore a highly conserved transcription factor in ExPEC strains, although strikingly it has been lost in many InPEC strains. A recent study showed that concurrent carriage of the LEE and *dsdCXA* was a rare occurrence in *E. coli*, with only 1.6% of 1,591 strains tested carrying both, suggesting a genetic 'incompatibility' between the two loci (Connolly *et al.*, 2015). Indeed, D-serine inhibits the formation of the T3SS, a key virulence determinant of EHEC and EPEC strains, providing a powerful selective pressure to lose *dsdCX* in these *E. coli* pathotypes (Connolly *et al.*, 2015).

DsdC is known to autoregulate itself and the *dsdXA* operon. However, any wider regulatory functions are currently unknown. This work therefore aimed to identify whether DsdC regulated genes beyond the *dsdCXA* locus and whether exposure to D-serine led to any global effects on regulation in two clinically important ExPEC strains, CFT073 and CE10. CFT073 and CE10 are two prototypic UPEC and NMEC strains respectively and were selected due to their carriage of the *dsdCXA* operon and thereby their ability to metabolise D-serine. Although they are both ExPEC, they utilise distinct mechanisms to infect and persist within the host.

CFT073 is a highly virulent UPEC strain that was isolated from a patient with acute pyelonephritis (Mobley *et al.*, 1990). Previous reports have indicated that D-serine may influence pathogenesis of CFT073, with a Δ*dsdA* mutant reportedly showing "hypercolonisation" and "hypermobility" in comparison to the WT (Roesch *et al.*, 2003). However, further research revealed these "hyper" phenotypes were due to an unrecognised secondary mutation in *rpoS*, rather than the Δ*dsdA* mutation (Hryckowian *et al.*, 2015). Nonetheless, D-serine remains an important factor for UPEC colonisation, with expression of *dsdA* and *dsdX* increasing 2.4-8.3-fold and 7.4-7.8-fold respectively in UPEC strain 83972 *in vivo* (Roos & Klemm, 2006).

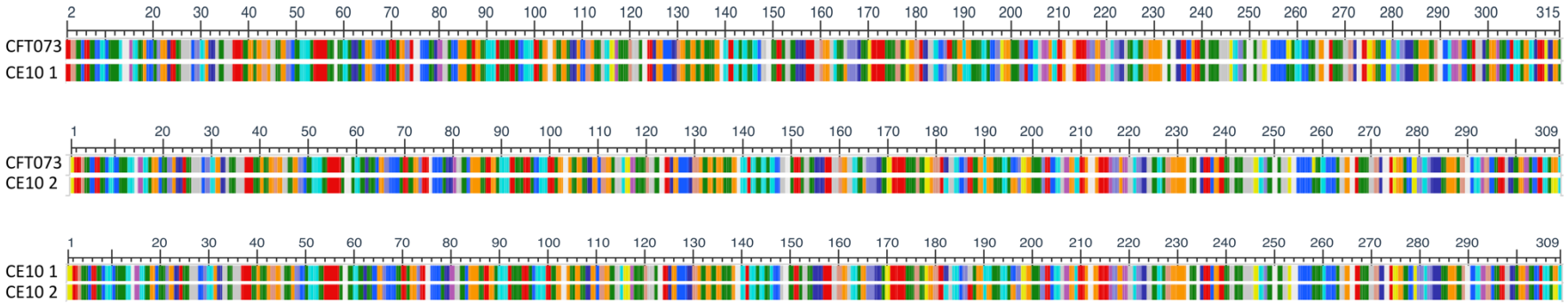
CE10 is a K1 NMEC strain that was isolated from the CSF of a patient with meningitis (Yao *et al.*, 2006). Like many K1 NMEC strains, CE10 carries two unlinked copies of the *dsdCXA* locus. It has been hypothesised that NMEC may be able to use D-serine as a carbon source for growth in the brain or as an environmental signal for gene expression (Moritz & Welch, 2006). Indeed, it has recently been shown that D-serine affects the transcriptome of CE10, causing differential expression of 55 genes and intriguingly repressing genes involved in acid tolerance (Connolly *et al.*, 2021). This work aimed to understand whether both copies of the NMEC CE10 DsdC were homologous in functionality. This chapter sought to address these questions by characterising the roles of CFT073 DsdC and CE10 DsdC1 and DsdC2. ChIP-Seq and RNA-Seq experiments were used to assess the global binding and the transcriptional response of CFT073 and CE10 DsdC in the presence and absence of D-serine.

3.2 DsdC and D-serine metabolism

3.2.1 Structural differences of DsdC

As with many NMEC K1 isolates, CE10 is interesting as it carries two copies of the *dsdCXA* locus. As it is yet unknown why CE10 carries two copies, the differences in amino acid structure of DsdC1 and DsdC2, as well as CFT073 DsdC were analysed.

A



B

```
CFT073_DsdC     MEPLREIRNRLLNGWQLSKLHTFEVAARHQSFALAAEELSLSPSAVSHRINQLEEELGIQ 60
CE10_DsdC1      MEPLREIRNRLLNGWQLSKLHTFEVAARHQSFALAAEELSLSPSAVSHRINQLEEELGIQ 60
CE10_DsdC2      MEPLREIRNRLLNGWQLSKLHTFEVAARHQSFALAAEELSLSPSAVSHRINQLEEELGIQ 60
                ************************************************************

CFT073_DsdC     LFVRSHRKVELTHEGKRVYWALKSSLDTLNQEILDIKNQELSGTLTLYSRPSIAQCWLVP 120
CE10_DsdC1      LFVRSHRKVELTHEGKRVYWALKSSLDTLNQEILDIKNQELSGTLTLYSRPSIAQCWLVP 120
CE10_DsdC2      LFVRSHRKVELTHEGKRVYWALKSSLDTLNQEILDIKNQELSGTLTLYSRPSIAQCWLVP 120
                ************************************************************

CFT073_DsdC     ALGDFTRRYPSISLTVLTGNDNVNLQRAGIDLAIYFDDAPSAQLAHHFLMDEEILPVCSP 180
CE10_DsdC1      ALGDFTRRYPSISLTVLTGNDNVNLQRAGIDLAIYFDDAPSAQLTHHFLMDEEILPVCSP 180
CE10_DsdC2      ALGDFTRRYPSISLTVLTGNDNVNLQRAGIDLAIYFDDAPSAQLTHHFLMDEEILPVCSP 180
                *********************************************:**************

CFT073_DsdC     EYAQRHDLTNTVINLRHCTLLHDRQAWSNDSGTDEWHSWAQHYAVNLPTSSGIGFDRSDL 240
CE10_DsdC1      EYAQRHALTNQVINLRHCTLLHDRQAWSNDSGTDEWHSWAQHYAVNLPTSSGIGFDRSDL 240
CE10_DsdC2      ENAQRHDLTNTVINLSHCTLLHDRQAWSNDSGTDEWHSWAQHYAVNLPTSSGIGFDRSDL 240
                * **** *** **** ********************************************

CFT073_DsdC     AVIAAMNHIGVAMGRKRMVQKRLASGELVAPFGDMTVKCHQHYYITTLPGRQWPKIEAFI 300
CE10_DsdC1      AVIAAMNHIGVAMGRKRLVQKRLASGELVAPFGDMTVKCHQHYYITTLPGRQWPKIEAFI 300
CE10_DsdC2      AVIAAMNHIGVAMGRKRLVQKRLASGELVAPFGDMTVKCHQHYYITTLPGRQWPKIEAFI 300
                *****************:******************************************

CFT073_DsdC     TWLREQVSQYECYTL   315
CE10_DsdC1      TWLREQVSQYECYTL   315
CE10_DsdC2      TWLREQVCQSCQYQ-   314
                *******.*   *
```

Figure 3-1 Multiple sequence alignments of the amino acid sequence of CFT073 and CE10 DsdC.

Figure 3-1 Multiple sequence alignments of the amino acid sequence of CFT073 and CE10 DsdC. The sequence for CFT073 DsdC was downloaded from NCBI Reference Sequence: NC_004431.1; protein ID WP_001327519.1. For CE10 DsdC1 and DsdC2, the sequences were downloaded from the GenBank file CP003034.1; protein ID AEQ13502.1 and AEQ14956.1 respectively. (A) The multiple sequence alignment of CFT073 DsdC vs CE10 DsdC1, CFT073 DsdC vs CE10 DsdC2, and CE10 DsdC1 vs CE10 DsdC2, using NCBI Protein BLAST with the standard parameters used. This was visualised using NCBI Multiple Sequence Alignment Viewer 1.20.0 with the colouring reflecting Rasmol Amino Acid colours. Each amino acid is rendered in a different colour. (B) The multiple sequence alignment of CFT073 DsdC, CE10 DsdC1 and CE10 DsdC2 was generated on Clustal Omega (EMBL-EBI). Grey shading indicates differences in amino acid sequence.

Figure 3-1 shows the multiple sequence alignments of CFT073 DsdC and CE10 DsdC1 and DsdC2. This analysis revealed that the amino acid sequence of the three DsdCs were highly conserved, with few differences between them. Indeed, a protein BLAST confirmed that DsdC was highly conserved with >98% percentage identity, as presented in Table 3-1.

Alignments	Query cover (%)	E-value	Percentage identity (%)
CFT073 DsdC against CE10 DsdC1	100	0	98.73
CFT073 DsdC against CE10 DsdC2	98	0	98.38
CE10 DsdC1 against CE10 DsdC2	98	0	98.38

Table 3-1 Amino acid sequence homology between the DsdCs. Data derived from the NCBI protein BLAST. Query cover is the percentage of the sequences aligned, E-value is the Expect value which describes the number of hits expected to be seen by chance, and percentage identity is the extent of which two amino acid sequences are similar.

However, there were a few differences between the amino acid sequences. Figure 3-1(B) indicated that there was a small cluster of differences between

codon 182 and 199, and differences towards the end of the gene. Using Pfam (EMBL-EBI), DsdC was modelled to visualise the predicted domains of a classical LTTR.

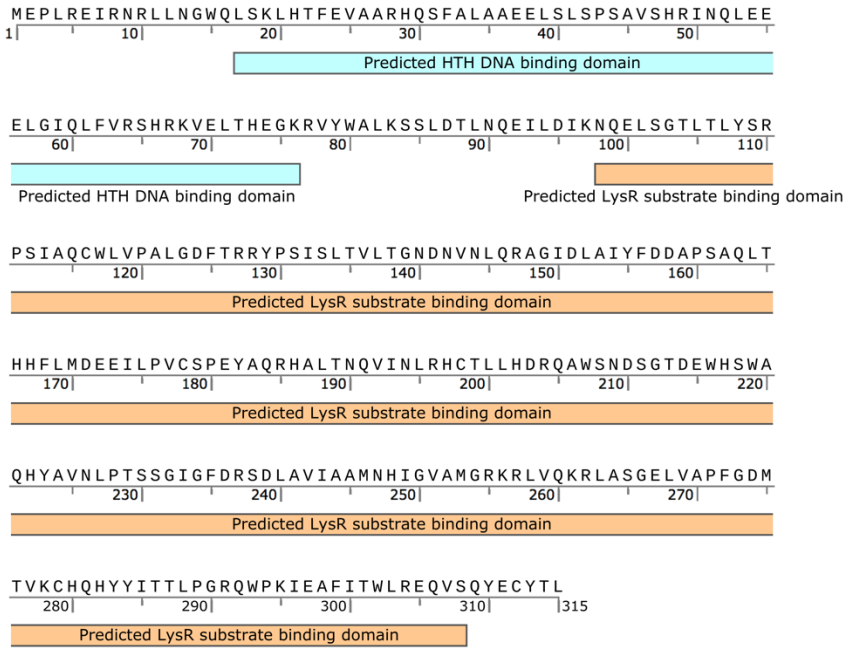


Figure 3-2 Visualisation of the predicted domains of DsdC. Using the CE10 DsdC1 amino acid sequence, UniProt entry A0A0E0V7F3, a Pfam model was generated to predict the HTH DNA binding domain and the predicted LysR substrate binding domain on CE10 DsdC1. E-values of these domains were 4.2e-16 and 3.5e-25 for the HTH and substrate binding domain respectively. E-values are based upon searching the Pfam-A family against UniProtKB using hmmsearch (Pfam version 34.0; EMBL-EBI). The data was visualised and exported from SnapGene.

Based upon the data obtained in Figure 3-2, it appeared that the differences in the amino acid sequence between CE10 DsdC1, DsdC2 and CFT073 DsdC were in the predicted substrate binding domain. As they were clustered closely together, the difference in amino acids has potentially caused a change in the tertiary structure, thus potentially altering the properties of the co-inducer binding pocket. Theoretically, this could enable a different substrate to bind to the TF. Table 3-2 shows the amino acid changes between CE10 DsdC1 and DsdC2.

CE10 DsdC1		CE10 DsdC2	
Amino acid	**Chemical property**	**Amino Acid**	**Chemical property**
Tyrosine (Y)	Polar uncharged (also hydrophobic)	Asparagine (N)	Polar uncharged
Alanine (A)	Hydrophobic	Aspartic acid (D)	Negative electronic charge
Glutamine (Q)	Polar uncharged	Threonine (T)	Polar uncharged
Arginine (R)	Positive electronic charge	Serine (S)	Polar uncharged

Table 3-2 Differences between the amino acid sequences in the substrate binding domain. The differences in amino acids and their chemical properties between CE10 DsdC1 and DsdC2 in the substrate binding domain.

The main differences highlighted in Table 3-2 were the changes between alanine (hydrophobic) to aspartic acid (negatively charged) and arginine (positively charged) to serine (uncharged). The chemical properties of amino acids can affect the folding of proteins. Hydrophobic amino acids are often buried within the core of the protein, whereas polar uncharged amino acids are often hydrophilic, clustering at the surface of a protein. The changes between the CE10 DsdC1 and DsdC2 structures may therefore result in a difference in the folding of the tertiary protein structure.

3.2.2 The role of DsdC in D-serine metabolism

Previous work by McFall and colleagues demonstrated that the transcriptional regulator DsdC was needed for D-serine metabolism in the laboratory *E. coli* strain K12 (Heincz *et al.*, 1984; Nørregaard-Madsen *et al.*, 1995). As CE10 contains two copies of this transcriptional regulator, both copies were assessed for functionality in D-serine metabolism. Furthermore, deletion of the regulator in both CFT073 and CE10 was assessed.

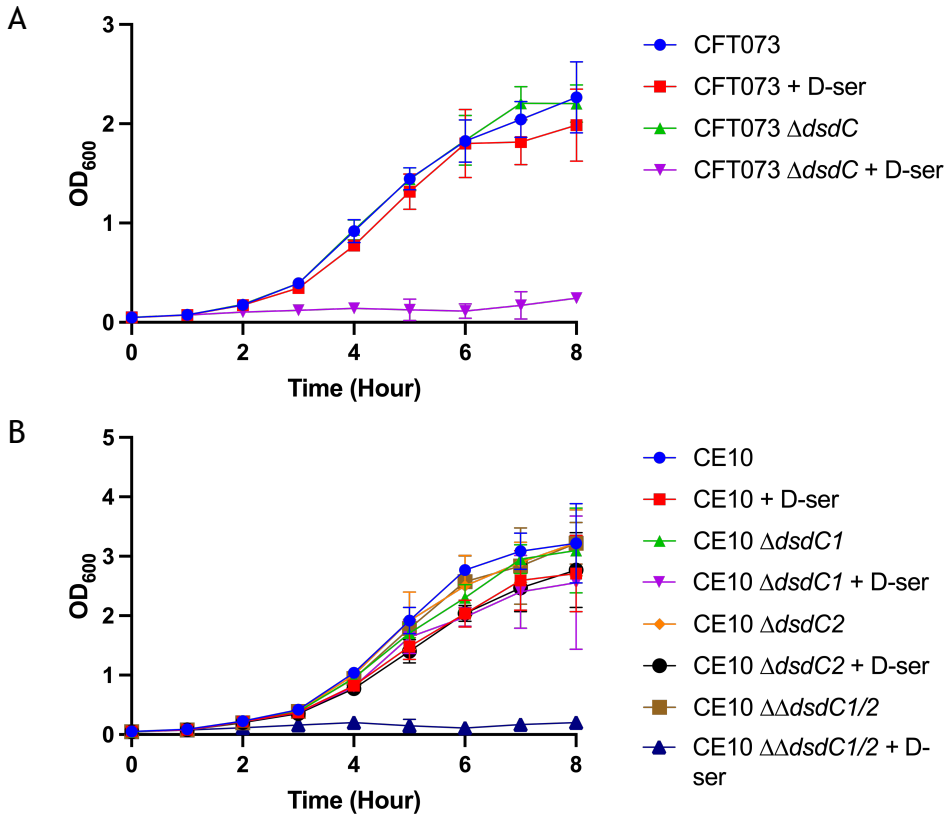


Figure 3-3 Growth profiles of WT and Δ*dsdC* in M9 minimal media supplemented with D-serine. Growth curves of CFT073 (A) and CE10 (B) in M9 minimal media supplemented with 1 mM of D-serine, measured as OD_{600} over time. Growth curves shown represent mean values of triplicate experiments with error bars indicating standard error of the mean (SEM).

Qualitative assessment of Figure 3-3(A) revealed that deletion of the regulator did not affect the growth of CFT073. This was the same for CE10 in Figure 3-3(B), where deletion of either the single regulators or the double regulator did not appear to impact growth. Figure 3-3(A) confirmed previous reports that DsdC was required for D-serine metabolism, with the CFT073 Δ*dsdC* mutant unable to grow in the presence of D-serine. Figure 3-3(B) showed that Δ*dsdC1* and Δ*dsdC2* could both metabolise D-serine, whereas the double ΔΔ*dsdC1/2* mutant could not. There appeared to be no difference in the growth between Δ*dsdC1* and Δ*dsdC2* when D-serine was supplemented into the media. These results indicated that both DsdC1 and DsdC2 were functionally redundant in enabling metabolism of D-serine, and it also appeared that one was not dominant to the other. This was also shown using minimal media sole carbon source plates in which D-serine was the only carbon source present (Fig. 3-4).

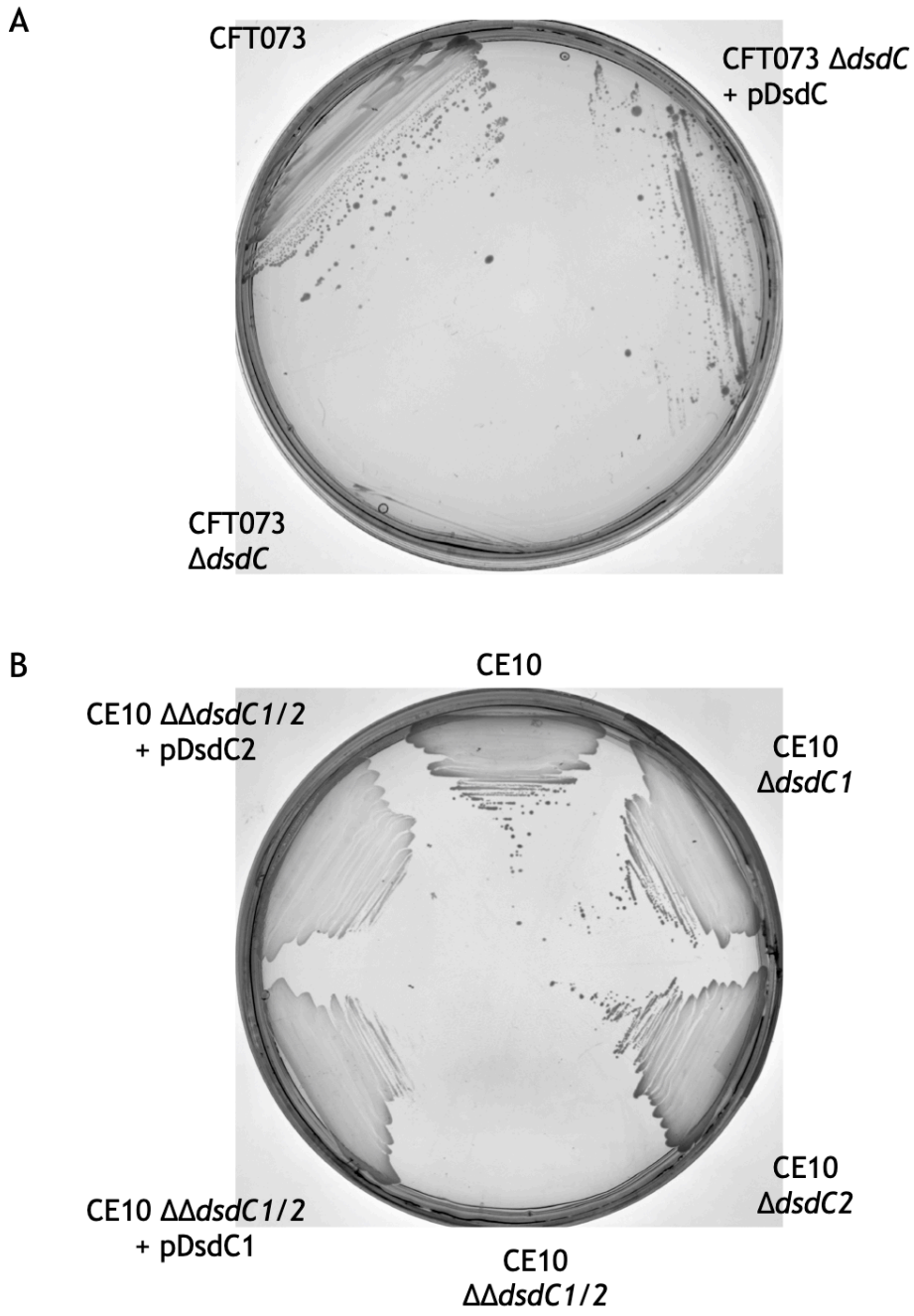


Figure 3-4 Growth profiles of WT and Δ*dsdC* on sole carbon source plates. Growth on minimal media plates where the sole carbon source was 10 mM D-serine. Depicting the characteristic growth profile of CFT073 (A) and CE10 (B) and isogenic mutants.

Figure 3-4(B) showed that both Δ*dsdC1* and Δ*dsdC2* could metabolise D-serine on the sole carbon source plates, with each appearing to grow as effectively as the other. However, as expected, the double ΔΔ*dsdC1/2* could not grow. Furthermore, the growth phenotype was restored when *dsdC1* and *dsdC2* were transformed into pACYC184 and complemented back into the double ΔΔ*dsdC1/2* mutant strain (Fig.3-5).

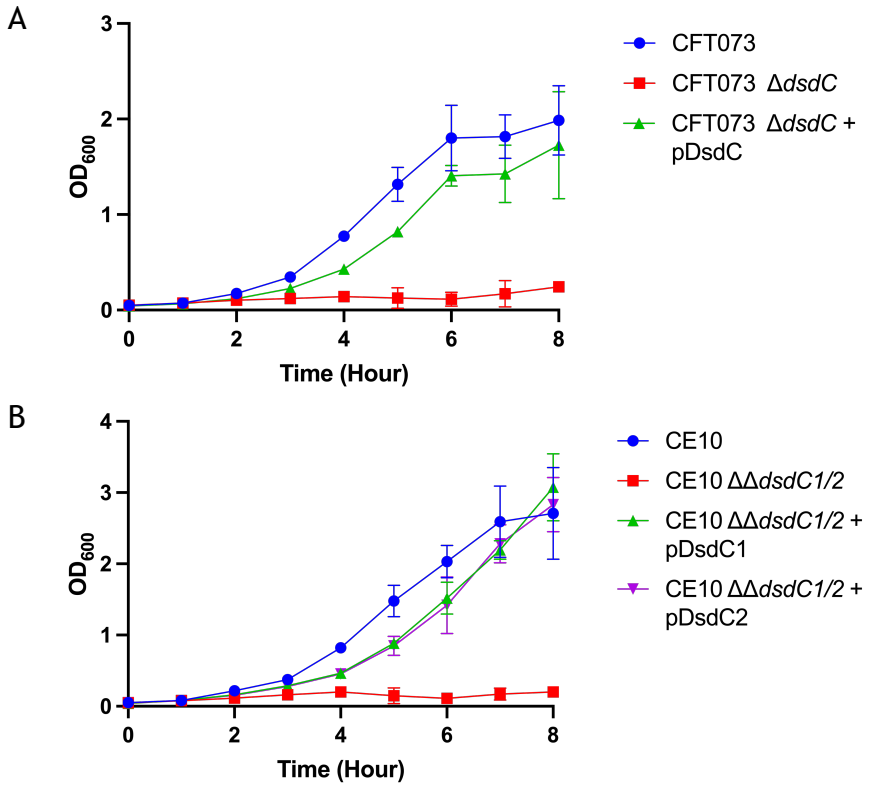


Figure 3-5 Growth profiles of WT, Δ*dsdC* and Δ*dsdC* + pDsdC in M9 minimal media supplemented with D-serine. Growth curves of CFT073 (A) and CE10 (B) in M9 minimal media supplemented with 1 mM of D-serine, measured as OD_{600} over time. Growth curves shown represent mean values of triplicate experiments with error bars indicating the SEM.

Qualitative assessment of Figure 3-5 revealed that when pDsdC was complemented back into the mutant, the growth phenotype was almost restored to WT levels in both CFT073 (A) and CE10 (B), although there was a slight growth lag. Nonetheless, pDsdC1 and pDsdC2 both appeared to grow comparably when complemented back into the double mutant ΔΔ*dsdC1/2*. This data correlated with previous reports that DsdC is essential in both CFT073 and CE10 for the metabolism of D-serine, with the complementation of pDsdC restoring growth. Furthermore, the results indicated that DsdC1 and DsdC2 appear to be functionally redundant with regard to D-serine metabolism.

3.3 Growth conditions and native expression of DsdC

3.3.1 Experimental design and growth conditions for the ChIP and RNA-Seq

Based on previous data from the Roe laboratory, it was established that the ChIP-Seq experiment would be performed under conditions matching an existing RNA-Seq experiment. Bacterial strains were grown in M9 minimal media for 3 hours at 37°C, 200 RPM. A subset was spiked with 1 mM D-serine, and all samples were grown for a further 2 hours at 37°C, 200 RPM. These experiments were designed to test the direct and indirect regulatory response of DsdC in relation to D-serine exposure. These conditions were chosen to limit the toxic effects that occurs within the Δ*dsdC* strains when D-serine is added to the samples at the start of the experiment, as shown in Figure 3-3.

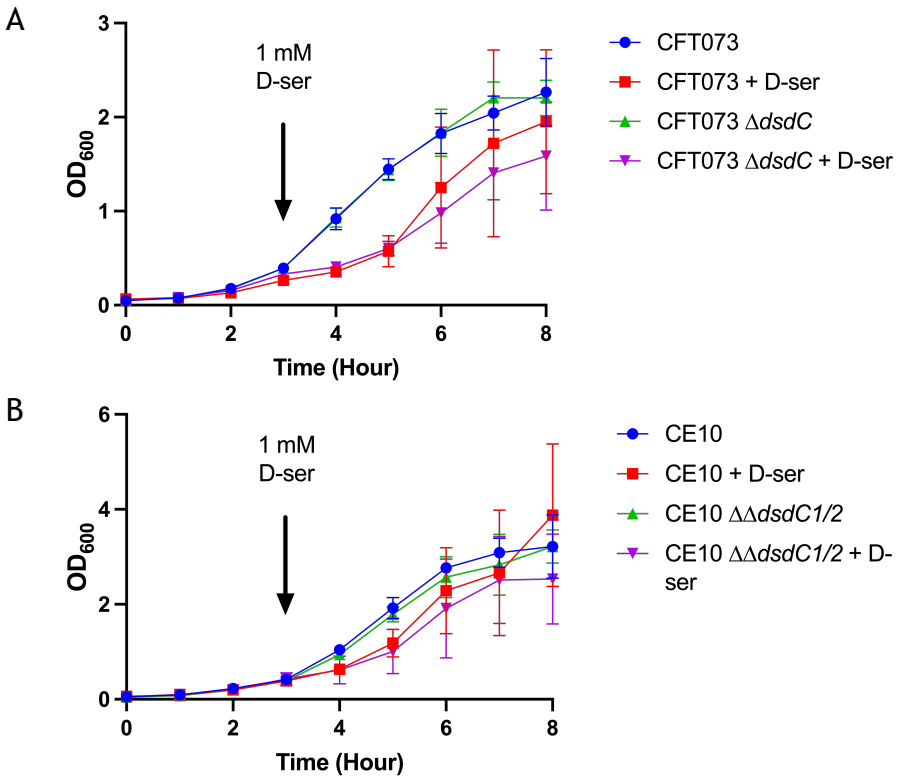


Figure 3-6 Growth profiles of WT and Δ*dsdC* in M9 minimal media spiked with D-serine. Growth curves of CFT073 (A) and CE10 (B) in M9 minimal media spiked with 1mM of D-serine at hour 3, measured as OD_{600} over time. Growth curves shown represent mean values of triplicate experiments with error bars indicating SEM.

Qualitative assessment of Figure 3-6 revealed when D-serine was spiked into CFT073 WT and Δ*dsdC* mutant at hour 3 (Fig. 3-6A), there was a slight growth defect. This growth defect was also present in CE10 (Fig. 3-6B), although not to the same extent. The growth in CE10 and ΔΔ*dsdC1/2* mutant was restored to that of the samples which had no additional D-serine by hour 6. CFT073 and CE10 were both sampled at hour 5, during the logarithmic growth phase of the bacteria, when cell uniformity is most likely.

3.3.2 Native expression of DsdC

To establish if DsdC was being expressed at the time point chosen for the ChIP-Seq, the native expression of DsdC was analysed for CFT073 and CE10, in the presence and absence of D-serine. All copies of DsdC were chromosomally FLAG-tagged to ensure only the native levels of expression were observed. These were then qualitatively assessed to ensure that the DsdC-FLAG tagged strains grew comparably to the WT (Fig. 3-7).

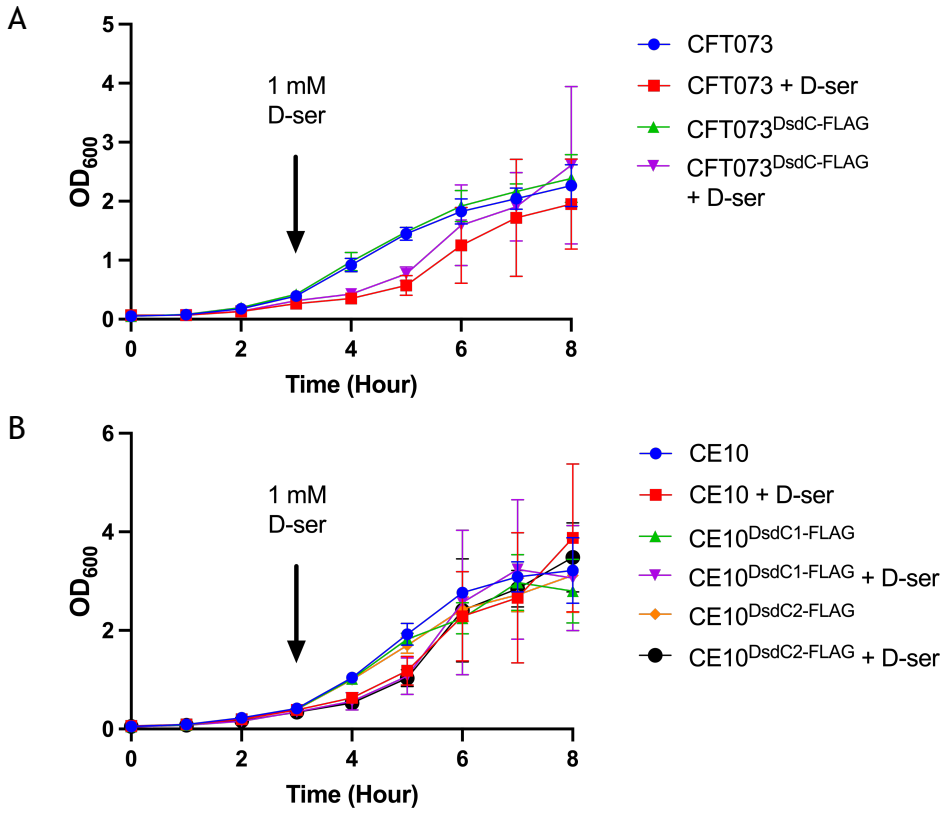


Figure 3-7 Growth profiles of WT and DsdC^{FLAG} mutants in M9 minimal media, spiked with D-serine. DsdC^{FLAG} tagged *E. coli* were assessed in M9 minimal media to ensure growth was similar to the non-FLAG tagged WT, measured as OD_{600} over time. This was performed with and without addition of 1 mM D-serine

at hour 3 for CFT073 (A) and CE10 (B). Growth curves shown represent mean values of triplicate experiments with error bars indicating SEM.

No qualitative differences were observed in growth between WT and DsdC^{FLAG} tagged strains (Fig. 3-7), indicating that engineering the FLAG tag to DsdC had had no impact upon growth. The native expression of DsdC was then analysed using SDS-PAGE and Western blot analysis (Fig. 3-8).

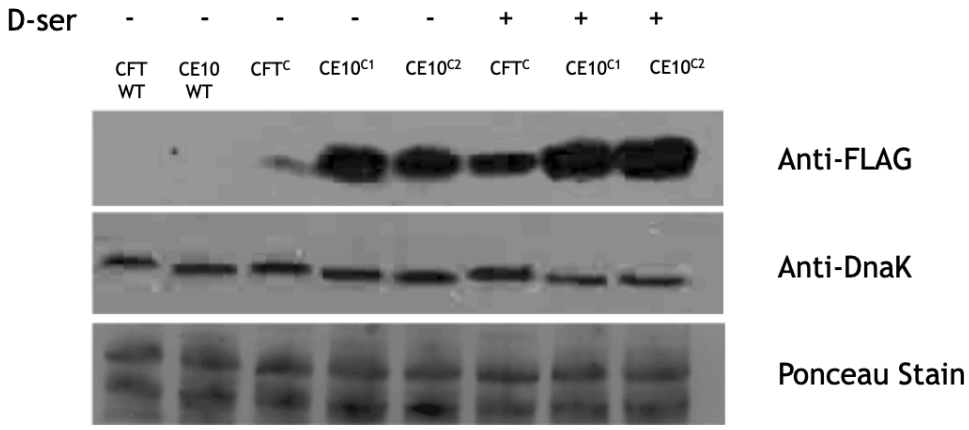


Figure 3-8 The native protein expression of DsdC in CFT073 and CE10. Western blot analysis of DsdC^{FLAG} expression from whole cell lysate of CFT073 and CE10 at the time point hour 5, in the presence and absence of D-serine, cultured in M9 minimal media. CFT073 and CE10 WT were used as negative controls. CFT^{C} represents CFT073$^{DsdC\text{-}FLAG}$, CE10^{C1} represents CE10$^{DsdC1\text{-}FLAG}$, and CE10^{C2} represents CE10$^{DsdC2\text{-}FLAG}$ This figure is representative of 3 separate experiments.

Western blot analysis revealed the relatively low abundance of CFT073 DsdC, in comparison to CE10 DsdC1 and DsdC2, in both the presence and absence of D-serine (Fig. 3-8). This finding potentially indicates that there are different regulatory mechanisms that control DsdC expression between the two strains. Western blot analysis also revealed that there was no difference in expression between DsdC1 and DsdC2, suggesting that there was no preferential favour to either DsdC1 or DsdC2 in the cell. When D-serine was spiked into the samples, it caused expression of all three DsdCs to increase, which correlates with previous findings. As a control, anti-DnaK antibody and Ponceau stain was used to ensure normal housekeeping expression and no differential loading of samples. In addition to investigating the native protein expression of DsdC at hour 5, the transcriptional levels of *dsdC* were also analysed.

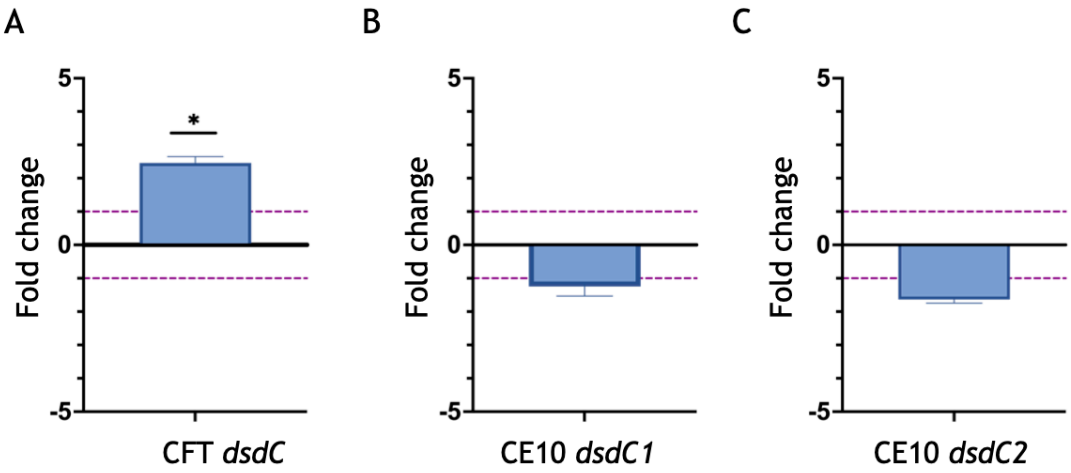


Figure 3-9 RT-qPCR fold change levels of *dsdC* in response to D-serine. RT-qPCR analysis of the transcriptional fold change levels of CFT073 *dsdC*, CE10 *dsdC1* and CE10 *dsdC2*, with * denoting significance of *p*-value ≤0.05. The purple dashed line indicates the baseline expression of *dsdC* at hour 5. (A) CFT073 *dsdC* with a fold change of 2.46 and a *p*-value of 0.00027. (B) CE10 *dsdC1* with a fold change of -1.24 and a *p*-value of 0.9. (C) CE10 *dsdC2* with a fold change of -1.63 and a *p*-value of 0.7.

Figure 3-9 indicated that for CE10 *dsdC1* and *dsdC2,* there was no significant difference in gene expression between the addition or absence of D-serine at hour 5, suggesting that *dsdC* expression had reverted to basal levels. Conversely, CFT073 *dsdC* was shown to be significantly upregulated when D-serine was spiked in (2.46-fold higher; *p*-value 0.0002). This suggests that potentially DsdC is regulated differently between the two strains, as both transcriptionally and translationally they function differently.

3.4 The global binding profile of CFT073 DsdC and CE10 DsdC1 and DsdC2

Until 2007, TF studies were limited to a small number of genetic locations, usually based upon prior knowledge of where a TF would be likely to be bound. However, with the advent of ChIP-Seq it is now possible to map protein-DNA interactions in an unbiased, high-throughput manner across the whole genome (Barski *et al.*, 2007; Johnson *et al.*, 2007; Mikkelsen *et al.*, 2007; Robertson *et al.*, 2007). To assess natively where DsdC was binding on the ExPEC genomes, $DsdC^{FLAG}$ tagged strains were genetically engineered and a ChIP-Seq experiment

performed. The conditions used for the ChIP-Seq experiment were performed to correlate with a previous RNA-Seq experiment and are described in Figure 3-10.

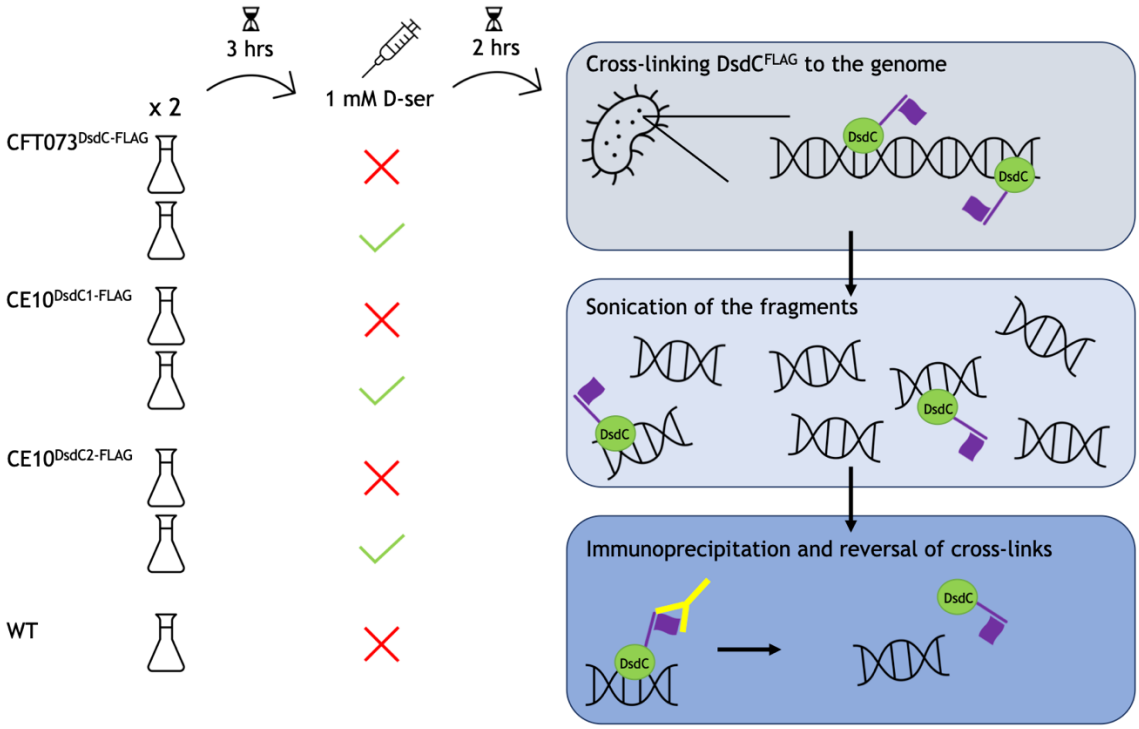


Figure 3-10 Graphical representation of the ChIP aspect of ChIP-Seq. 2 biological replicates of the DsdC^{FLAG} tagged strains were grown in M9 minimal media for 3 hours, 200 RPM at 37°C. 1 mM of D-serine was added to a subset of the samples and the samples were incubated for a further 2 hours. Proteins were then cross-linked to the DNA using formaldehyde and the DNA was sheared into 200-300 bp fragments by sonication. The DNA fragments bound by DsdC were immunoprecipitated out and the cross-links were reversed. The isolated DNA was then sent for Illumina sequencing. A non-FLAG tagged WT was used as a mock ChIP control.

Once the chromatin was immunoprecipitated and cross-links reversed, the samples were sent for library preparation and Illumina Next Generation sequencing at Glasgow Polyomics, performed by Julia Galbraith. The reads were quality controlled (QC) checked, uploaded onto CLC Genomics Workbench 7 and mapped against the CFT073 and CE10 reference genomes; NCBI accession numbers NC_004431.1 and CP003034 respectively. The Transcription Factor ChIP-Seq tool was then used to analyse the genomic coverage of the reads,

identifying a set of positive regions with very apparent peaks. The mock ChIP experiment was used as a negative example for the software. Once the positive and negative regions were identified, the software builds a filter which was used to identify other genomic regions, of which the read coverage profile matches the characteristic peak shape (Fig. 3-11). Figure 3-12 shows the peak shapes generated for CFT073$^{DsdC\ FLAG}$, CE10$^{DsdC1\text{-}FLAG}$ and CE10$^{DsdC2\text{-}FLAG}$ tagged samples.

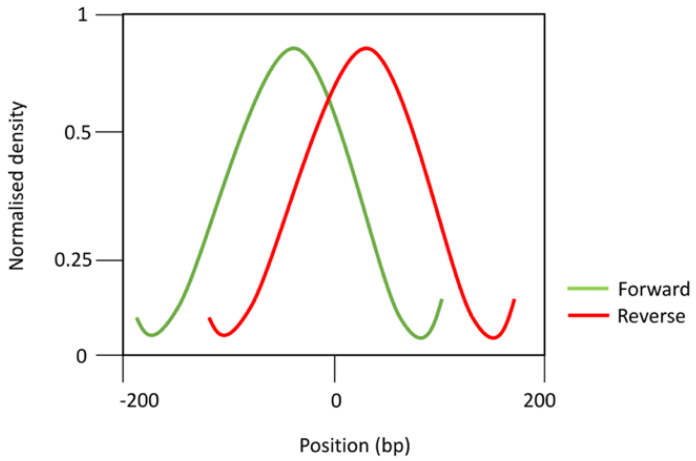


Figure 3-11 Canonical ChIP-Seq peak shape. The canonical peak shape of a transcription factor based upon the Hotelling Observer filter used in CLC Genomic Workbench 7 (QIAGEN). The green line indicates the forward read and the red line indicates the reverse read. Figure adapted from CLCBio (QIAGEN).

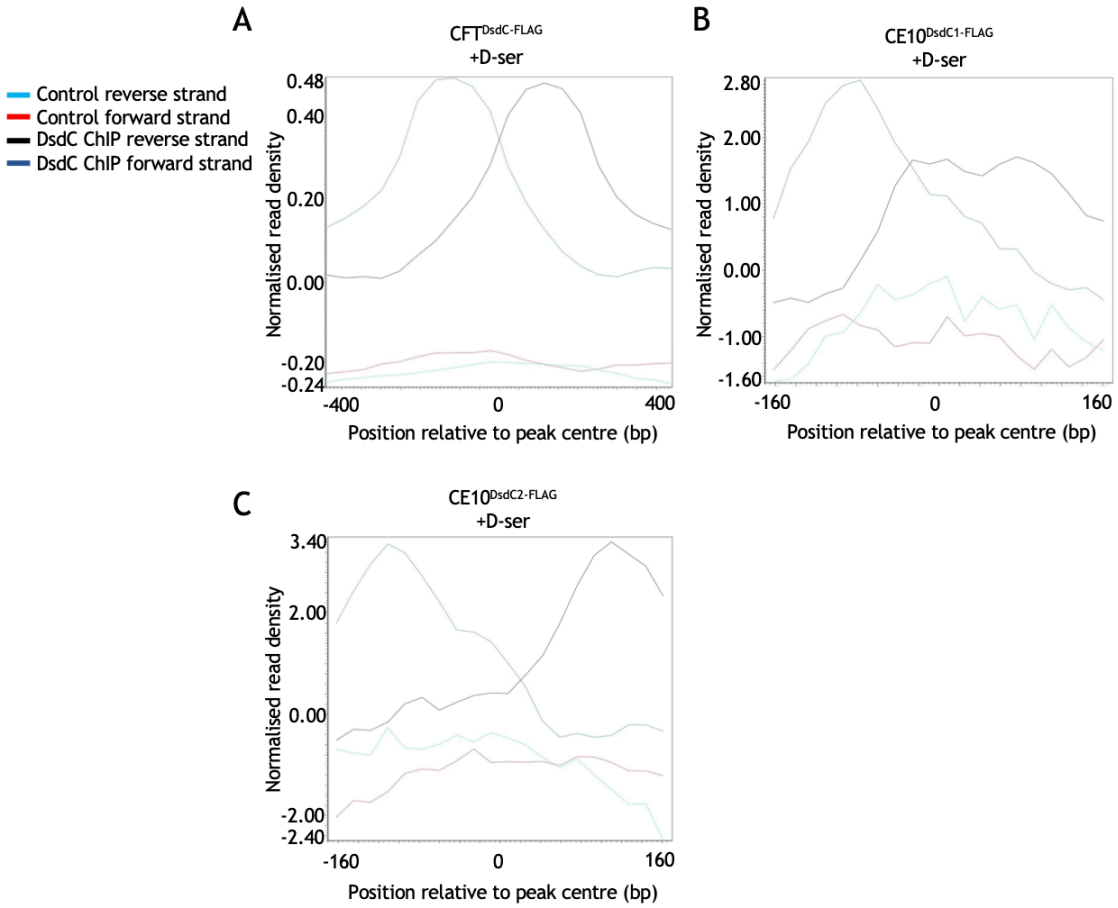


Figure 3-12 Peak shape filter score for CFT073$^{DsdC\text{-}FLAG}$, CE10$^{DsdC1\text{-}FLAG}$ and CE10$^{DsdC2\text{-}FLAG}$. The blue and black lines represent the ChIP forward and reverse strands respectively, whilst the red and aqua lines represent the mock ChIP control forward and reverse sample respectively. This shape should give the classical bimodal peak that genuine ChIP peaks must adhere to. (A) Peak shape filter score for CFT073$^{DsdC\text{-}FLAG}$ with D-serine; (B) CE10$^{DsdC1\text{-}FLAG}$ with D-serine; (C) CE10$^{DsdC2\text{-}FLAG}$ with D-serine. This data was obtained from CLC Genomics Workbench 7 (QIAGEN).

The CFT073 peak shape filter scores observed in Figure 3-12 represented the canonical TF peak shape that genuine TF peaks must adhere to. However, the CE10 peak shape score appeared to be slightly broader in frame. Subsequently, the peaks were manually checked to look at the binding profiles. Interestingly, two distinct peak shapes emerged from the ChIP-Seq data (Fig. 3-13). The first peak shape (Fig. 3-13A/C/E) was the canonical TF peak shape, whereas the second peak shape was broader and covered larger regions of the genome (Fig. 3-13B/D/F).

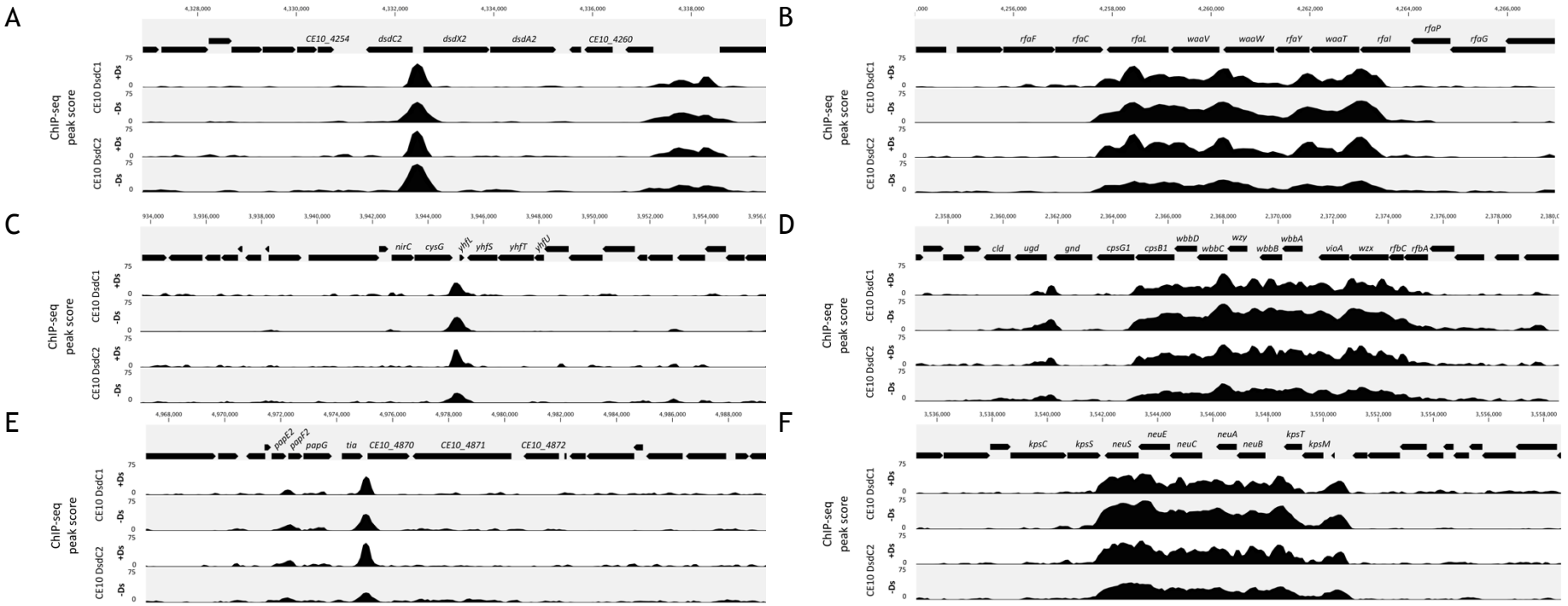


Figure 3-13 Expanded view of ChIP-Seq peaks from CE10. An expanded view of the ChIP-Seq peaks from CE10$^{DsdC1\text{-}FLAG}$ and CE10$^{DsdC2\text{-}FLAG}$ highlighting the differences in binding patterns. (A, C and E) show the canonical ChIP TF binding peak. (B, D and F) show a broader peak encompassing small sections of the genome.

Due to the high number of false positives associated with ChIP-Seq, peaks called by the software were manually checked until there were 10 in a row that did not conform to the canonical TF shape. Table 3-3 shows the number of ChIP-Seq peaks called by the CLC software, and after manual inspection the number of peaks that were deemed to be "true" binding peaks (data listed in Supplementary Table S1-6).

Sample	Number of statistically significant peaks called (*p*-value ≤0.05)	Number of "true" ChIP-Seq peaks
CFT073$^{DsdC\text{-}FLAG}$	375	88
CFT073$^{DsdC\text{-}FLAG}$ + 1mM D-serine	1,115	129
CE10$^{DsdC1\text{-}FLAG}$	7,001	95
CE10$^{DsdC1\text{-}FLAG}$ + 1mM D-serine	5,724	217
CE10$^{DsdC2\text{-}FLAG}$	4,820	140
CE10$^{DsdC2\text{-}FLAG}$ + 1mM D-serine	5,498	177

Table 3-3 Analysis of the ChIP-Seq peaks called from the CLC ChIP-Seq analysis software. The number of statistically significant peaks (*p*-value ≤0.05) called using CLC and a summary of the potential "true" direct binding sites of DsdC that was established after manual inspection of the ChIP-Seq data.

As Table 3-3 shows, the CLC software predicted >250 peaks for CFT073$^{DsdC\text{-}FLAG}$and >1,000 peaks for CE10$^{DsdC1\text{-}FLAG}$ and CE10$^{DsdC2\text{-}FLAG}$. Upon manual inspection however, many of these were deemed false positives and the number of "true" DsdC binding peaks ranged between 88 and 217 peaks. These results indicated that DsdC was more than just the transcriptional regulator for D-serine metabolism and that it was significantly binding to multiple regions across the genome.

3.4.1 Genome wide binding of DsdC

Using the ChIP-Seq data, the global binding profile of DsdC was overlayed onto the CFT073 and CE10 genome (Fig. 3-14).

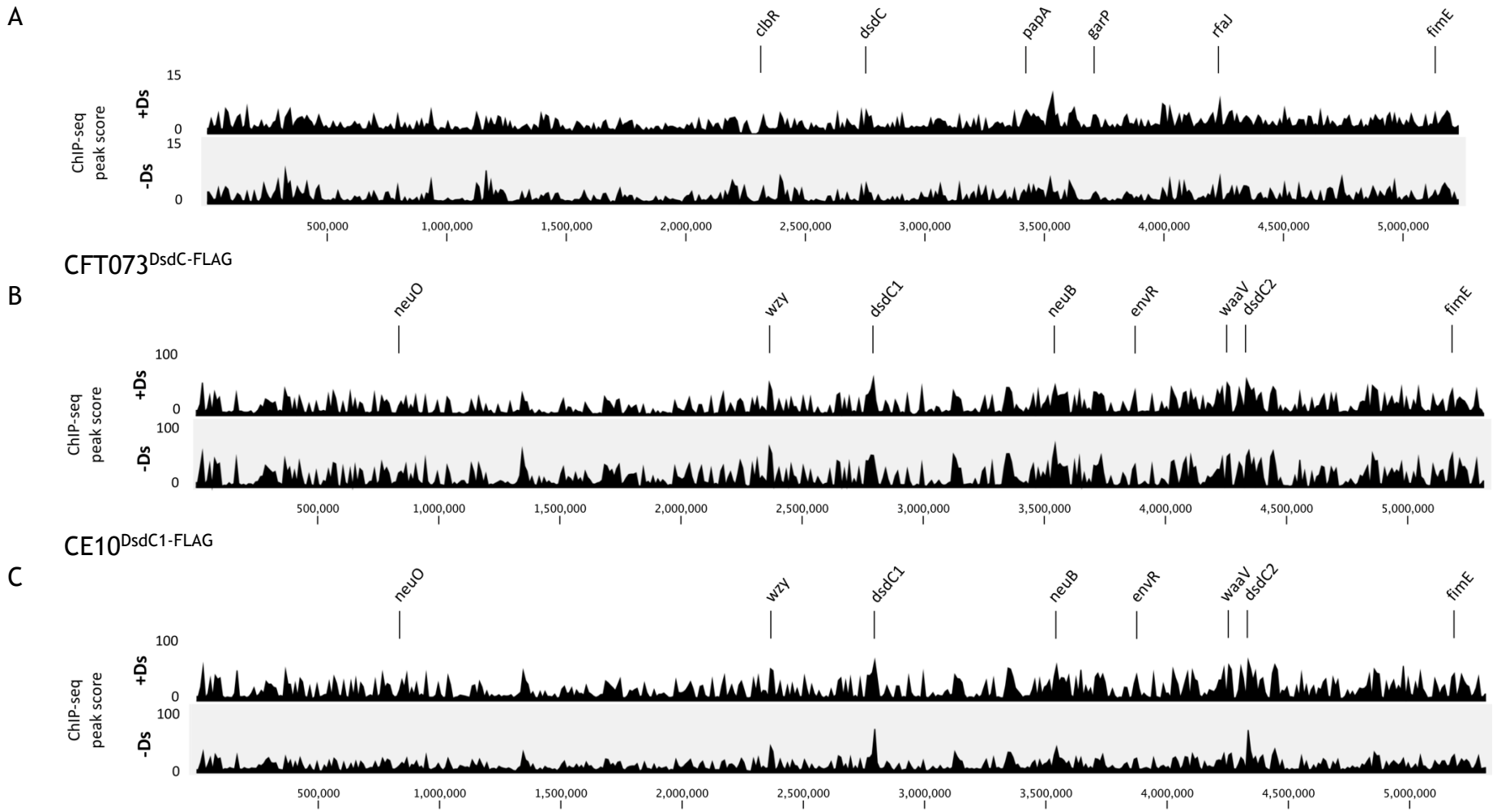


Figure 3-14 The global binding profile of CFT073$^{DsdC\text{-}FLAG}$, CE10$^{DsdC1\text{-}FLAG}$ and CE10$^{DsdC2\text{-}FLAG}$.

Figure 3-14 The global binding profile of CFT073$^{DsdC\text{-}FLAG}$, CE10$^{DsdC1\text{-}FLAG}$ and CE10$^{DsdC2\text{-}FLAG}$. The global binding profile of the genomic regions bound by DsdC. Significant peaks (*p*-value ≤0.05) of interest are labelled above each track map in black. (A) CFT073$^{DsdC\text{-}FLAG}$ ChIP-Seq peak score in the presence and absence of D-ser; (B) CE10$^{DsdC1\text{-}FLAG}$ ChIP-Seq peak score in the presence and absence of D-ser; (C) CE10$^{DsdC2\text{-}FLAG}$ ChIP-Seq peak score in the presence and absence of D-ser.

Analysis of Figure 3-14 indicated that the DsdC peaks were distributed across the whole genome, and that DsdC was not solely binding in a single region. These results indicated that DsdC could potentially be a global regulator, and was not just acting locally on the *dsdCXA* locus. The track maps of the ChIP-Seq peak scores (Fig. 3-14A-C) highlighted how similar the DsdC binding profiles were in the presence or absence of D-serine. Indeed, upon further analysis of the ChIP-Seq data, it was revealed that there was no difference in binding patterns between presence and absence of D-serine in each of the data sets. Further, there was no difference in binding patterns observed between CE10 DsdC1 and DsdC2. Peaks that appeared to be unique to DsdC1 were in fact present in the DsdC2 sample, and vice versa, but had not made the "canonical" TF shape cut-off. It is therefore likely that there was no real difference in binding sites between the two homologous TFs. Instead, any differences observed were likely to be due to experimental artefacts. In contrast, there were differences observed between the binding sites for DsdC in CE10 and CFT073, indicating for the first time that there was strain-specific regulation mediated by DsdC.

3.4.2 Functional characterisation of the ChIP-Seq hits and strain-specific binding of DsdC

Functional analysis and characterisation was carried out on the ChIP-Seq hits. Genes with the 5' end closest to the nearest associated ChIP binding peak were assumed to be the gene regulated by DsdC (Fig. 3-15).

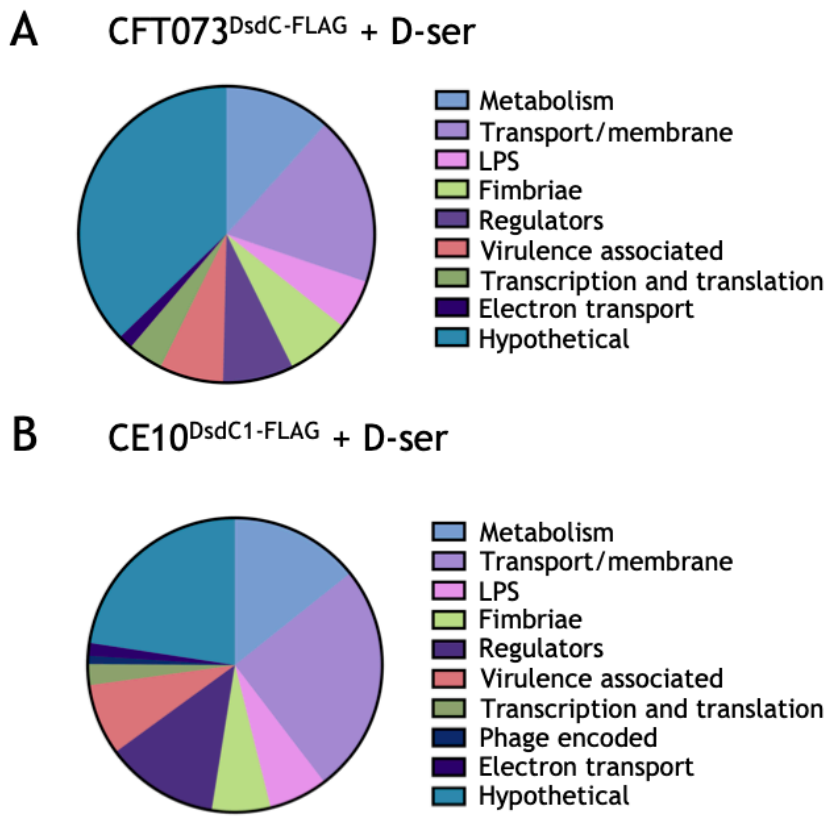


Figure 3-15 Functionality of the genes predicted as direct binding targets of DsdC. Genes with 5' ends closest to the nearest associated ChIP binding peak were functionally characterised using UniProtKB GO functional groupings (Bateman et al., 2021). Functional groups were broken down into broad categories such as metabolism and virulence associated genes for (A) CFT073$^{DsdC\text{-}FLAG}$ + D-ser and (B) CE10$^{DsdC1\text{-}FLAG}$ + D-ser.

Functional analysis of the genes predicted to be regulated by CFT073 DsdC revealed great insight into the potential roles of the TF. Indeed, 18% of the genes (24/129) that were bound were involved in transport or were membrane associated. These included genes such as *pitB*, a phosphate transporter; c5298, a hexuronate transporter; and *garP*, a galactarate transporter. Intriguingly, DsdC also bound to several virulence associated regions in CFT073: 7% of the genes (9/129) were associated with fimbriae production, including *papA*, *fimE*, and *fimA*, and 5% of the genes (7/129) were associated with LPS biosynthesis including *rfaJ*, *waaL* and *waaV*. These were segregated from the general 'virulence associated' functional grouping to highlight the large abundance of genes within these categories. The remaining functional groups included genes involved in metabolism (15/129), other regulators (10/129), virulence associated (9/129), transcription and translation (5/129), electron transport (2/129), and

uncharacterised (48/129). These results illustrated that DsdC binds to a wide range of genes that are involved in multiple different functional mechanisms in CFT073, further indicating the possibility of DsdC as a global regulator.

As the ChIP binding profiles had revealed no differences in binding pattern between CE10 DsdC1 and DsdC2, the data set of CE10$^{DsdC1\text{-}FLAG}$ + D-ser was used as a representative of the group. Gene ontology (GO) analysis revealed the largest functional group bound by DsdC was transport and membrane associated genes which accounted for 25% of the targets (55/217). These included genes such as *fepE*, a ferric enterobactin transporter; *fucP*, an L-fucose transporter; and *feoA*, a ferrous iron transporter. Further, DsdC was also shown to bind to genes involved in LPS and capsular biosynthesis (14/217), with targets such as *neuO* and *neuB*. The K1 capsule plays an essential role in NMEC pathogenicity and is encoded on the *kps* operon. This operon is not present in CFT073, potentially therefore alluding to strain-specific regulation by DsdC of a key virulence mechanism in CE10. The remaining functional groups included genes involved in metabolism (31/217), fimbriae (14/217), other regulators (27/217), virulence associated (17/217), transcription and translation (5/217), phage encoded (2/217), electron transport (3/217) and uncharacterised (49/217). These results all indicated a global role of DsdC in binding to the CE10 genome.

Interestingly, the global binding and functional analysis of the ChIP-Seq results had revealed strain-specific binding of DsdC in CFT073 and CE10. Indeed, highly conserved TFs have recently been shown to be "repurposed", uniquely controlling separate gene sets in different bacteria and tailoring gene regulation to suit strain-specific lifestyles (reviewed in O'Boyle *et al.*, 2020). A comparison of the ChIP-Seq binding peaks between CFT073$^{DsdC\text{-}FLAG}$ and CE10$^{DsdC1\text{-}FLAG}$, both in the presence of D-serine, was therefore made (Fig. 3-16; Supplementary Table 7). NCBI BLAST was used to map gene sequences to the relevant reference genome for gene homology.

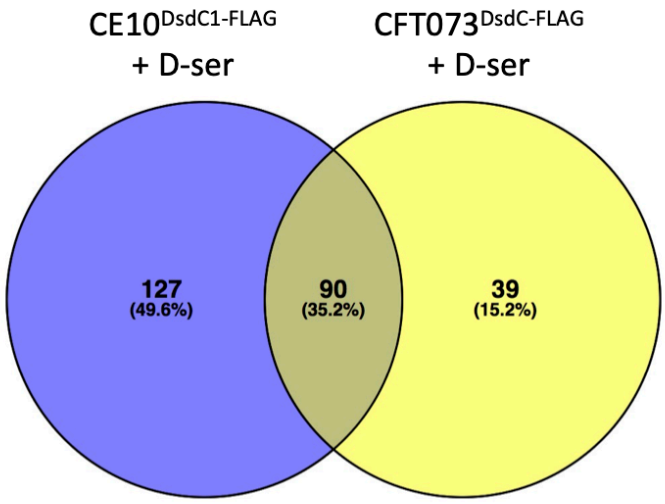


Figure 3-16 Strain-specific binding of DsdC. A Venn diagram highlighting the difference in binding sites between CFT073$^{DsdC\text{-}FLAG}$ and CE10$^{DsdC1\text{-}FLAG}$, in the presence of D-serine. NCBI BLAST was used to assess homology of genes. Diagram made on Venny 2.1.0 (Oliveros, 2007).

Figure 3-16 showed that 35% of targets (90 genes) predicted to be bound by DsdC were common to both CFT073 and CE10. Functional characterisation of the commonly bound genes revealed functions involved in both core processes of the cell such as metabolism and transport but also virulence associated genes, specifically LPS biosynthesis genes and putative fimbrial genes (Fig. 3-17).

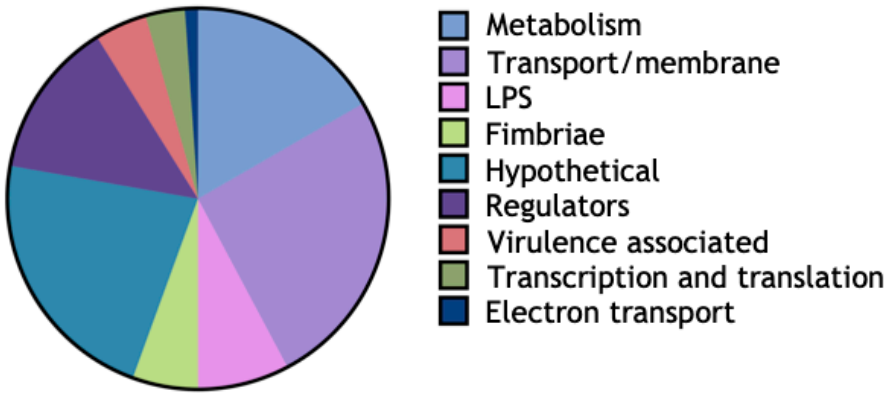


Figure 3-17 Functional analysis of common binding peaks between CFT073$^{DsdC\text{-}FLAG}$ and CE10$^{DsdC1\text{-}FLAG}$. A graphical representation of the shared binding peaks and their associated gene functionality between CFT073$^{DsdC\text{-}FLAG}$ and CE10$^{DsdC1\text{-}FLAG}$.

Notably however, there were differences between the data sets, with CE10 DsdC1 uniquely binding to 127 regions of the genome, and CFT073 DsdC uniquely binding to 39, indicating that there was strain-specific binding by DsdC (Fig. 3-16). Of the 39 putatively bound genes that were unique to the CFT073 data set, 26 of these were specific to CFT073 and had no homology in CE10. The 13 other genes all had homology in CE10 and contained a peak in the CE10 data set; however, the peaks did not meet the peak threshold criteria. Therefore, only the 26 genes that were completely unique to CFT073 will be discussed (Fig. 3-18A).

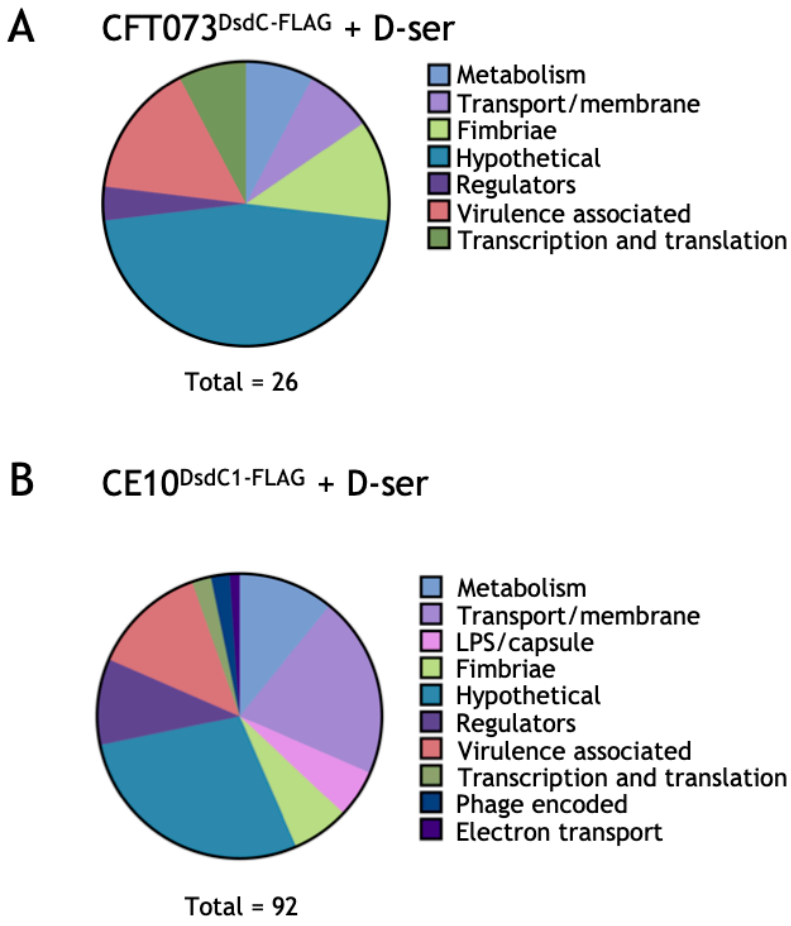


Figure 3-18 Strain-specific functionality of DsdC binding sites. A graphical representation of the unique binding peaks and their associated gene ontology in (A) CFT073$^{DsdC\text{-}FLAG}$ and (B) CE10$^{DsdC1\text{-}FLAG}$.

Functional analysis revealed virulence associated genes were one of the largest functional groups in the unique CFT073 DsdC binding data set (4/26) and included the genes *tssB* and *yheE*. These genes are involved in the type 6 secretion system (T6SS) and a putative type 2 secretion system (T2SS)

respectively. Both of these mechanisms are utilised by bacteria to secrete effector molecules into the host or neighbouring bacteria and are key virulence mechanisms. Furthermore, it was shown that DsdC also bound uniquely to *clbR*, the regulator of colibactin biosynthesis. Colibactin is a bacterial genotoxin produced by UPEC and has recently been associated with colorectal cancer in humans. The other genes uniquely bound by CFT073 DsdC were found to be involved in metabolism (2/26), transport and membrane associated (2/26), fimbriae (3/26), transcription and translation (2/26) and uncharacterised (12/26). These results indicated that DsdC bound uniquely to multiple genes, and strikingly to several virulence associated genes, indicating that DsdC has potentially been tailored in CFT073 strains.

Of the 127 genes that were unique to the CE10 data set, 67 of these were specific to CE10, 25 had homology in CFT073 but no peak, and 35 had homology in CFT073 and had a peak but did not meet the peak threshold criteria (Fig. 3-16). Therefore, only the 92 genes that appeared to be completely unique to CE10 will be discussed (Fig. 3-18B). Interestingly, many of the potential genes bound by CE10 DsdC were also found to be associated with virulence. Indeed, capsular and LPS genes (5/92), fimbriae genes (6/92) and general virulence associated genes (12/92) accounted for 26% of the unique CE10 hits. Notably, genes involved in the NMEC putative secondary T3SS ETT2, were shown to be bound by DsdC. This virulence factor is not carried by CFT073 (Ren et al., 2004) and again could indicate a role for DsdC in strain-specific regulation. The other genes found to be bound uniquely by CE10 DsdC were involved in metabolism (10/92), transport and membrane associated (19/92), other regulators (9/92), transcription and translation (2/92), phage encoded (2/92), electron transport (1/92), and uncharacterised (26/92). These results all indicate that DsdC has been tailored in ExPEC to mediate different regulatory pathways specific to individual strain requirements.

3.4.3 Genomic context of DsdC binding

Where on the genome a TF binds to, in relation to the 5' end of a gene, can reveal a lot about the dynamics of the TF. Indeed, the canonical TF model is that TFs bind at the 5' region of a gene, usually in intergenic non-coding regions. The genomic context of DsdC binding were therefore analysed (Fig. 3-19).

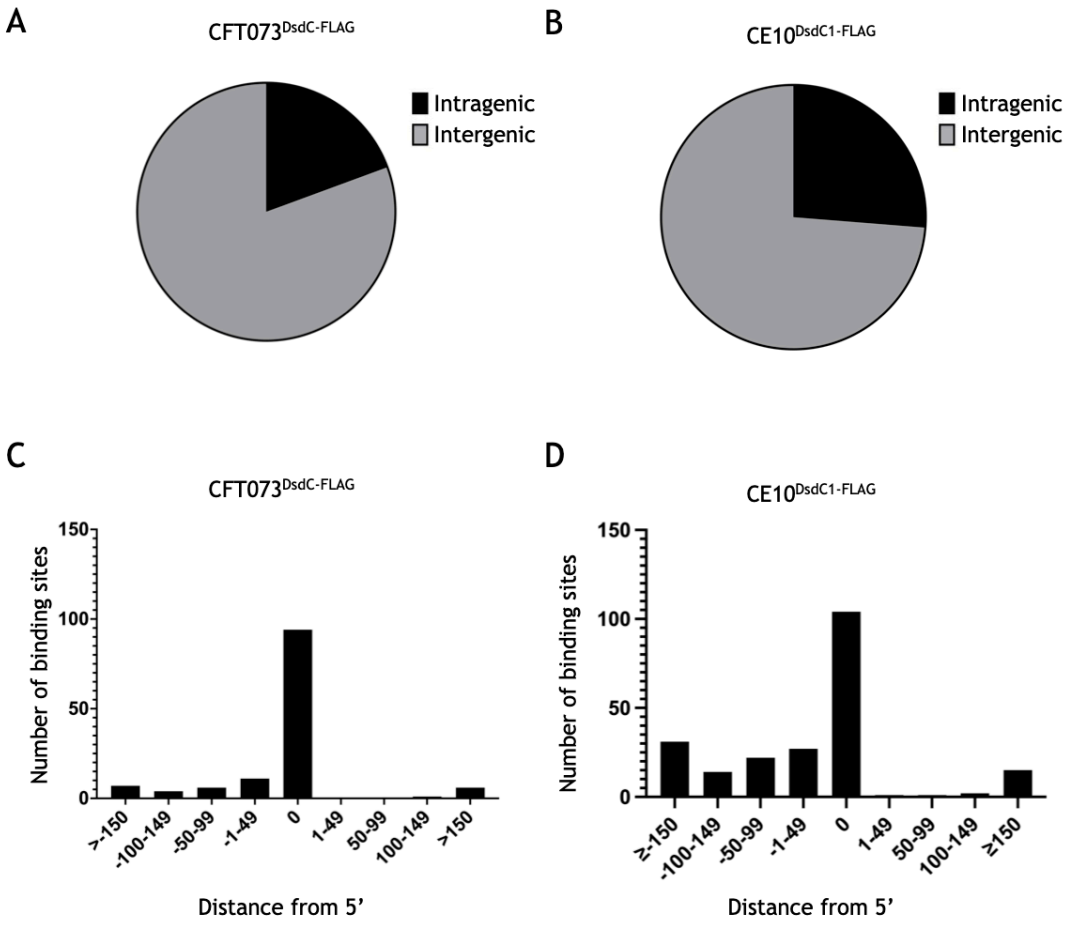


Figure 3-19 Genomic context of ChIP-Seq binding hits. The position of the ChIP-Seq binding peak in terms of intergenic and intragenic regions of the genome for (A) CFT073$^{DsdC\text{-}FLAG}$ + D-ser and (B) CE10$^{DsdC1\text{-}FLAG}$ + D-ser. The distance (bp) of the ChIP-Seq binding peak position in relation to the start of the 5' end of the gene for (C) CFT073$^{DsdC\text{-}FLAG}$ + D-ser and (D) CE10$^{DsdC1\text{-}FLAG}$ + D-ser. Raw data for figures are in Supplementary Table S1-6.

Figure 3-19(A-B) shows the majority of the CFT073 and CE10 DsdC ChIP peaks were binding to intergenic regions of the genome, 81% and 74% respectively. However, there were peaks that were binding in ORFs, with 19% of the CFT073 peaks (25/129) and 26% of the CE10 peaks (57/217) binding in intragenic regions. The distance of the binding peak, in relation to the 5' end of the gene, was also analysed (Fig. 3-19C-D). As would be expected of a canonical TF, DsdC binds mainly at the 5' end of the gene, with most peaks being in the 0 bp 'bin' away from the 5' end. However, there was a vast range in binding site distance for DsdC, with 7 and 31 binding sites over 150 bp downstream of the 5' end for CFT073 and CE10 respectively. A large proportion of the peaks were downstream to the 5' end of the gene with 122 out of 127 CFT073 DsdC peaks and 198 out of

217 CE10 DsdC peaks fitting this description. These results all indicated that DsdC was binding as a classical TF, with the majority of peaks being intergenic and near the TSS of the gene.

Further analysis of the ChIP-Seq data could be performed in relation to the DsdC binding motif. A binding motif is the common consensus sequence that a TF uses to recognise and bind to DNA. The generally accepted LTTR binding motif consists of a 13 bp sequence T-N_{11}-A, although this can vary in base pair composition and length (Goethals *et al.*, 1992; Maddocks & Oyston, 2008). Although ChIP-Seq maps DNA-TF interactions at a high resolution, this resolution is not high enough to map the TF binding motif (Wade, 2015), and therefore the MEME-Suite software was used to generate the DsdC binding motif (Bailey *et al.*, 2009). The 60 most significant predicted peaks of $CE10^{DsdC1\text{-}FLAG}$ + D-ser were used to generate the binding motif. The standard parameters were used in the MEME-Suite software with advanced options stating a motif minimum width of 11 bp and maximum width of 27 bp. The software was to run until 10 motifs had been found. However, no significant binding motifs under these parameters were generated. As there had been two types of binding patterns observed in the ChIP-Seq data analysis, the search was refined using only the top 20 predicted binding regions that had a canonical peak shape. The same advanced options were used and three statistically significant predicted binding motifs were generated (Fig. 3-20).

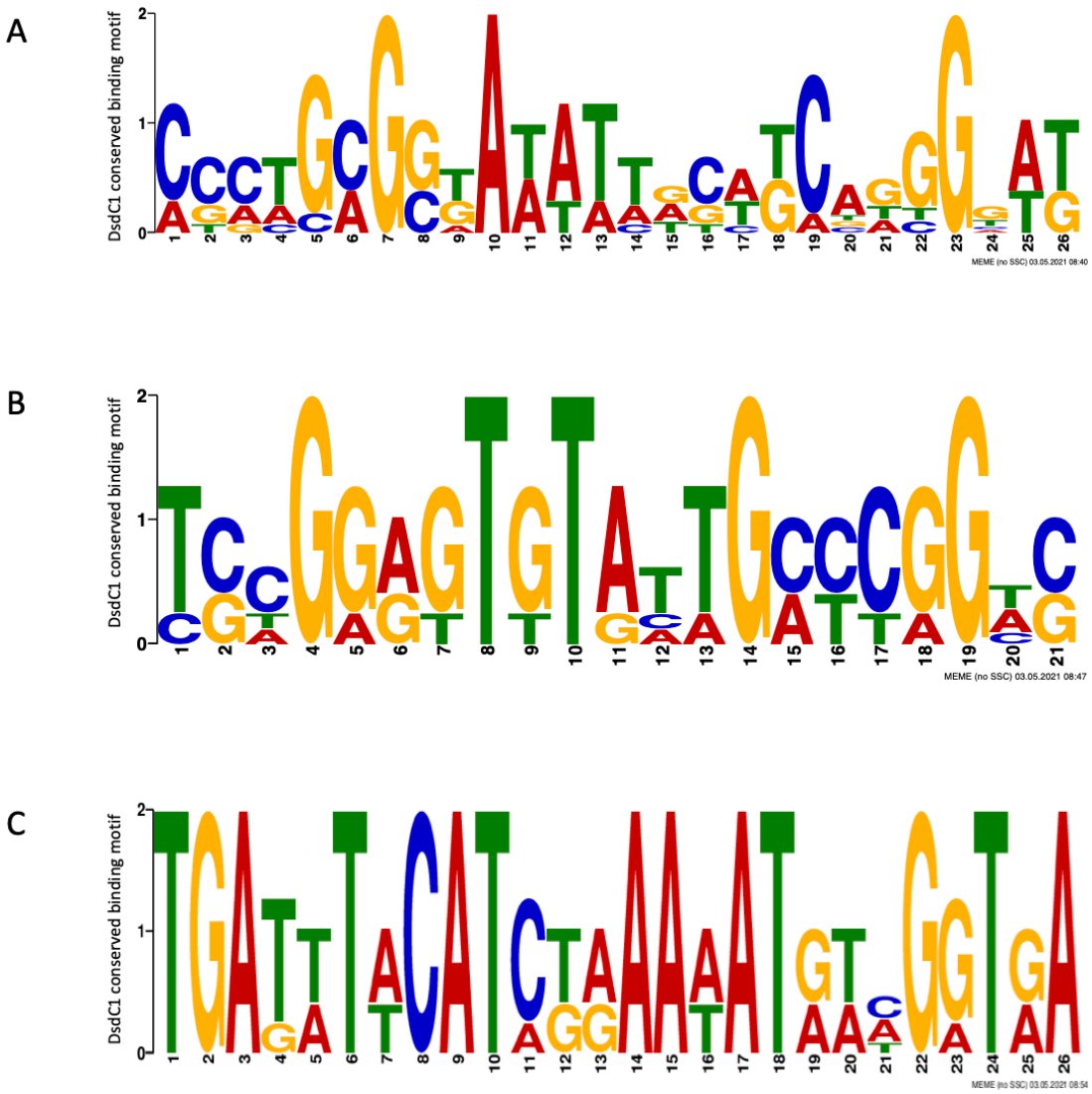


Figure 3-20 Predicted DsdC binding motifs. Predicted DsdC binding motifs that were generated from the CE10$^{DsdC1\text{-}FLAG}$ + D-ser dataset using MEME-Suite. These binding motifs were statistically significant with E-values of (A) 5.5e-007, (B) 1.2e-002, and (C) 3.3e-002.

Figure 3-20 presents three predicted DsdC binding motifs generated from the ChIP-Seq data. The LTTR binding box is the consensus sequence T-N_{11}-A, however, none of the motifs generated fit that consensus sequence. Conversely, Figure 3-20(A) predicts two T-N_{11}-A sequences joined together and Figure 3-20(C) predicts a consensus sequence of T-N_{24}-A. Although these predictations are statistically significant, DNase I footprinting should be used to experimentally validate the binding motif. Together, the data presented here indicated that DsdC is a global TF, that binds in intergenic and intragenic regions across the genome.

3.5 Transcriptomic regulon of CFT073 DsdC and CE10 DsdC1/2

Although DsdC was shown to bind globally across the genome in the ChIP-Seq binding data, this did not demonstrate transcription of downstream genes. Indeed, direct binding of a TF does not necessarily equate to a transcriptional response. Global transcriptome profiling by RNA-Seq can identify the direct and indirect regulon of a given TF, revealing novel networks and pathways. Therefore, to determine whether DsdC was playing a direct regulatory role in CFT073 and CE10, a transcriptomics experiment was carried out, comparing the RNA levels between WT and mutant, in the presence and absence of D-serine. This was performed by Dr Nicky O'Boyle, as described in 2.3.7, with the conditions used represented in Figure 3-21.

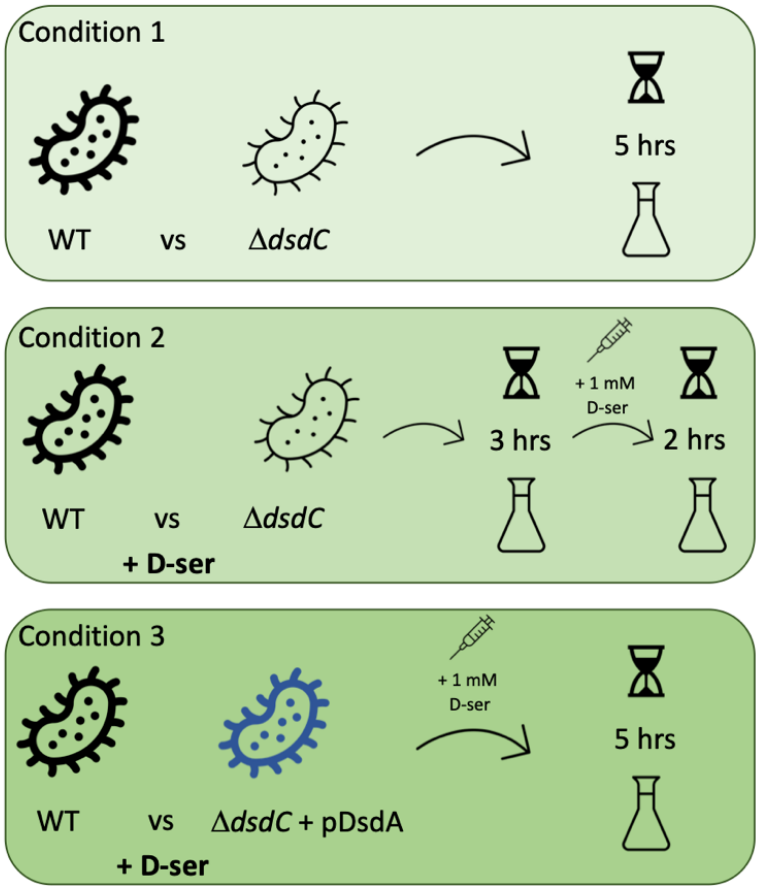


Figure 3-21 Conditions used in the RNA-Seq experiment. A graphical representation of the conditions used in the RNA-Seq experiment in CFT073 and CE10. Condition 1 examined the transcriptional changes between "WT vs.

Δ*dsdC*" (or ΔΔ*dsdC1/2* for CE10). Strains were grown for 5 hours in M9 minimal media. Condition 2 examined the transcriptional changes between "$WT^{D\text{-}ser}$ vs. Δ*dsdC*$^{D\text{-}ser}$" (or ΔΔ*dsdC1/2*$^{D\text{-}ser}$ for CE10). Strains were grown for 3 hours in M9 minimal media, 1 mM of D-serine was added, and strains were grown for a further 2 hours. Condition 3 examined the transcriptional changes between "$WT^{D\text{-}ser}$ vs. Δ*dsdC* + $pDsdA^{D\text{-}ser}$" (or ΔΔ*dsdC1/2* + $pDsdA^{D\text{-}ser}$ for CE10). Strains were grown for 5 hours in M9 minimal media with 1 mM D-serine added in from the start.

Figure 3-21 shows a graphical representation of the three conditions used in the RNA-Seq experiment. All conditions were performed in M9 minimal media, at 37ºC, 200 RPM. Condition 1 examined the transcriptional changes between "WT vs. Δ*dsdC*". This pairwise comparison sought to reveal any roles of the TF DsdC in the absence of D-serine. Condition 2 examined the transcriptional changes between "$WT^{D\text{-}ser}$ vs. Δ*dsdC*$^{D\text{-}ser}$". This condition was chosen to show the role of DsdC in the presence of D-serine. The samples were spiked with 1 mM D-serine at hour 3, to minimise any toxic side effects of D-serine on the mutant as described in 3.2.2 and 3.3.1. Further, as there was the potential that D-serine accumulation, even after spiking at hour 3, could cause detrimental effects upon the mutant and affect changes in gene regulation, pDsdA was complemented into the Δ*dsdC* mutant (condition 3). As DsdA is the D-serine deaminase, this experimental condition would reveal the transcriptional changes between "$WT^{D\text{-}ser}$ vs. Δ*dsdC*$^{D\text{-}ser}$", whilst theoretically minimising any potential D-serine accumulation side effects.

3.5.1 DsdC differentially regulates genes in the presence and absence of D-serine

The reads from the RNA-Seq experiment were quality assessed and mapped to the CFT073 and CE10 genome and plasmids (NCBI accession number: NC_004431.1, CP003034, CP003035, CP003036, CP003037, CP003038) using CLC Genomics Workbench 7. Differential expression was assessed using the empirical analysis of differential expression (EdgeR) with genes displaying absolute fold changes of ≥ 1.5; ≤ -1.5 and having a false-discovery rate corrected *p*-value of ≤0.05 being considered. Pairwise comparisons were done using the three

conditions stated above (Fig. 3-21). This analysis was performed by Dr Nicky O'Boyle prior to this PhD commencing. Table 3-4 shows a summary of the total number of differentially expressed genes (DEGs). Full differential expression data is available on ENA under the accession numbers detailed in Chapter 2. Figures 3-22 and 3-23 are volcano plots of the CFT073 and CE10 transcriptomic data respectively.

RNA-Seq pairwise comparison	**Total no. of differentially expressed genes (DEGs)**	**No. of genes upregulated**	**No. of genes downregulated**
CFT073 WT vs. Δ*dsdC* (condition 1)	47	29	18
CFT073 $WT^{D\text{-}ser}$ vs. Δ*dsdC*$^{D\text{-}ser}$ (condition 2)	1067	542	525
CFT073 $WT^{D\text{-}ser}$ vs. Δ*dsdC* + pDsdA$^{D\text{-}ser}$ (condition 3)	175	162	12
CE10 WT vs. ΔΔ*dsdC1/2* (condition 1)	82	42	40
CE10 $WT^{D\text{-}ser}$ vs. ΔΔ*dsdC1/2*$^{D\text{-}ser}$ (condition 2)	436	287	149
CE10 WT + D-ser vs. ΔΔ*dsdC1/2* + pDsdA1/2$^{D\text{-}ser}$ (condition 3)	552	300	252

Table 3-4 Summary of DEGs from the RNA-Seq experiment. A summary of the total number of DEGs from each pairwise comparison, and subdivided into number of DEGs upregulated and downregulated. Genes were classed as DEGs if they displayed absolute fold changes of ≥ 1.5; ≤ -1.5 and had a false-discovery rate corrected *p*-value of ≤0.05.

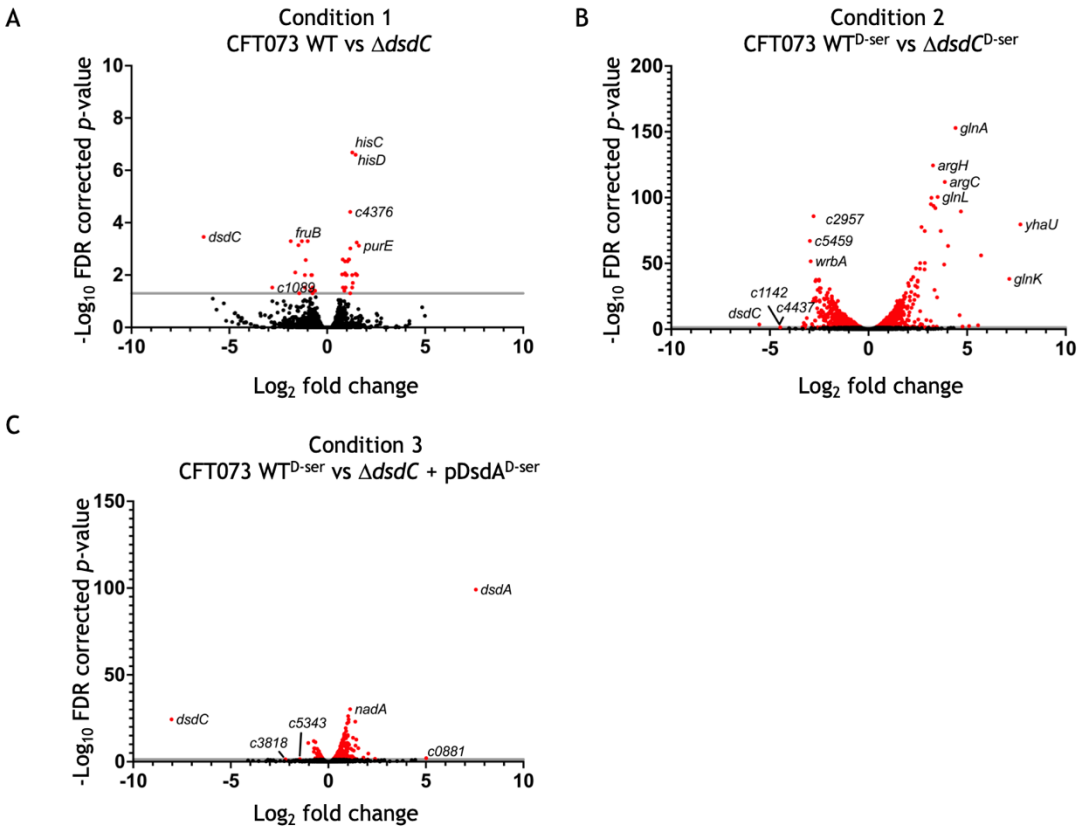


Figure 3-22 Volcano plots of CFT073 RNA-Seq transcriptome data. This data displays the Log_2 fold change in genes for the pairwise comparisons between (A) Δ*dsdC* compared to the WT; (B) Δ*dsdC* compared to the WT in the presence of 1 mM D-serine; (C) Δ*dsdC* + pDsdA compared to the WT in the presence of 1 mM D-serine. Significant DEG (*p*-value ≤0.05) are displayed in red, with the grey line indicating the *p*-value cut off for DEGs.

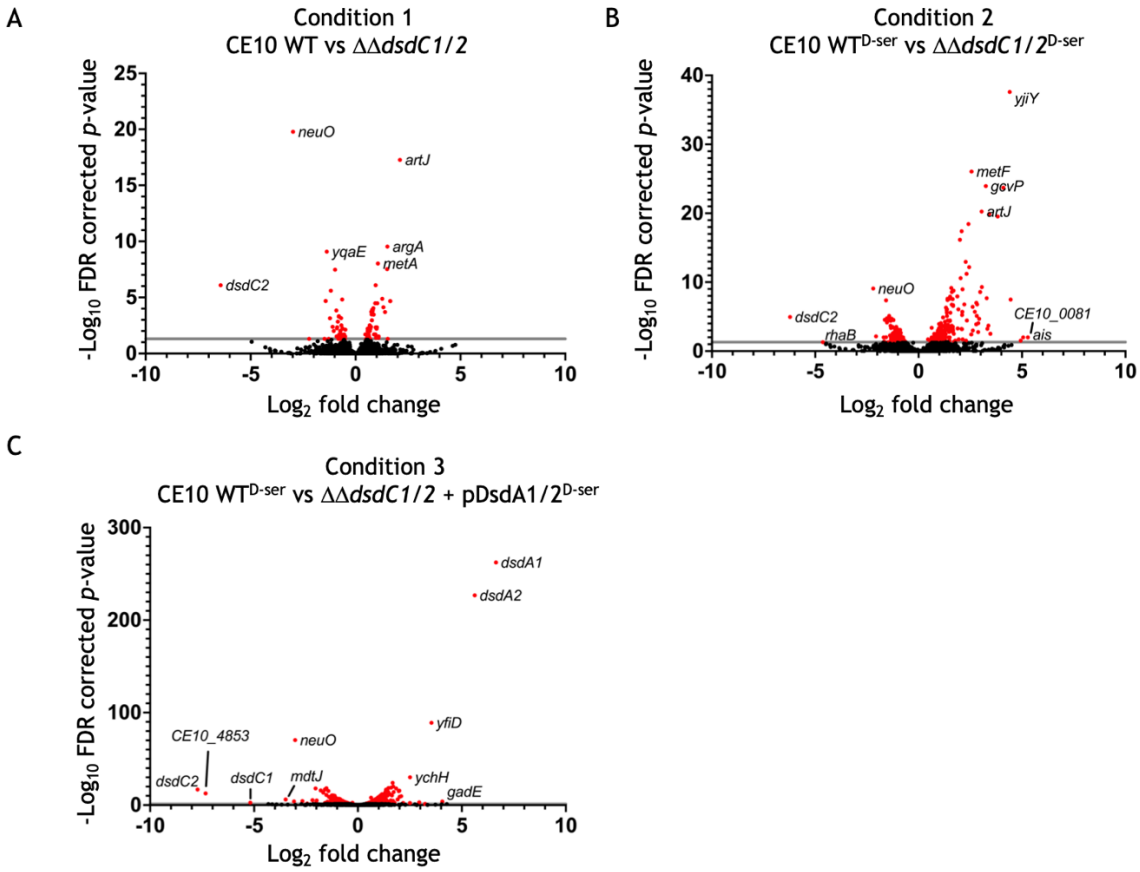


Figure 3-23 Volcano plots of CE10 RNA-Seq transcriptome data. This data displays the Log_2 fold change in genes for the pairwise comparisons between (A) Δ*dsdC1/2* compared to the WT; (B) Δ*dsdC1/2* compared to the WT in the presence of 1 mM D-serine; (C) Δ*dsdC1/2* + pDsdA1/2 compared to the WT in the presence of 1 mM D-serine. Significant DEG (*p*-value ≤0.05) are displayed in red, with the grey line indicating the *p*-value cut off for DEGs.

Table 3-4 displays the total number of DEGs from each pairwise comparison (visualised in the volcano plots; Fig. 3-22 and 3-23). Condition 1, "WT vs. Δ*dsdC*", yielded the least amount of DEGs for both CFT073 and CE10, 47 and 82 respectively. For CFT073, the fold changes of DEGs were relatively low, with all but 5 of the DEGs having fold changes ranging between >3 and <-3. The most upregulated DEG, *purE*, had a 3.07-fold change (*p*-value 7.61E-04), and the two most downregulated DEGs bar *dsdC*, *c1089* and *fruB*, had a -6.97 (*p*-value 0.03) and -3.64-fold change (*p*-value 5.08E-04) (Fig. 3-22A). The low fold change was also reflected in the CE10 data, in which all but 5 of the DEGs had fold changes between >3 and <-3. The most upregulated DEG, *artJ*, was 4.32-fold higher (*p*-value 5.31E-18) and the most downregulated DEG bar *dsdC2*, *neuO*, was -7.9-fold lower (*p*-value 1.66E-20) (Fig. 3-23A). Interestingly *dsdC1* was not a DEG in the CE10 condition 1 dataset. These results indicated that, in the absence of D-serine, DsdC is playing a limited role in gene expression in both CFT073 and CE10.

Conversely, when D-serine was added to the experiment (condition 2), the number of DEGs increased vastly, to 1,067 and 436 in CFT073 and CE10 respectively (Table 3-4; Fig. 3-22B and 3-23B). Furthermore, in condition 2, the fold changes observed were also more pronounced. In the CFT073 "$WT^{D\text{-}ser}$ vs. Δ$dsdC^{D\text{-}ser}$", the two most up-regulated genes, *yhaU* and *glnK*, had fold changes of 208.42 (*p*-value 2.76E-80) and 141.49 (*p*-value 6.33E-39), and the two most downregulated genes bar *dsdC*, *c4437* and *c1142*, had fold changes of -22.35 (*p*-value 0.04) and -22.18 (*p*-value 0.04). Indeed, 246 DEGs had fold changes ranging between >3 and <-3. This increase in fold change levels was also reflected in the "CE10 $WT^{D\text{-}ser}$ vs. ΔΔ$dsdC1/2^{D\text{-}ser}$" transcriptomic comparison. The two most upregulated genes, *ais* and *CE10_0081*, had fold changes of 38.93 (*p*-value 0.01) and 33.97 (*p*-value 0.01) and the two most downregulated genes bar *dsdC2*, *rhaB* and *neuO*, had fold changes of -24.76 (*p*-value 0.05) and -4.57 (*p*-value 8.56E-10). These results indicated that in the presence of D-serine there is a greater transcriptional effect of DsdC in both CFT073 and CE10. Indeed, the differences in number of DEGs was substantial between absence and presence of D-serine. To rule out any potential adverse side effects that may have arisen from D-serine accumulation in the Δ$dsdC^{D\text{-}ser}$ mutant, and therefore

any potential changes to the transcriptome as a consequence, *dsdA* was transformed into pACYC184 and complemented into Δ*dsdC* (condition 3).

Notably, there was a pronounced difference in the number of genes differentially expressed between the CFT073 "WT$^{D\text{-}ser}$ vs. Δ*dsdC* + pDsdA$^{D\text{-}ser}$" data set, in comparison to "WT$^{D\text{-}ser}$ vs. Δ*dsdC*$^{D\text{-}ser}$"; the number of DEGs dropped from 1,067 to 175. This result indicated that the majority of DEGs present in the "WT$^{D\text{-}ser}$ vs. Δ*dsdC*$^{D\text{-}ser}$" pairwise comparison was likely due to D-serine accumulation and toxicity within the cells. The degree of fold change was also less pronounced with only 7 genes, including *dsdA* and *dsdC* having fold changes of >3 and <-3. The two most significant upregulated genes bar *dsdA*, *yjiY* and *c0881* had fold changes of 41.57 (*p*-value 0) and 32.38 (*p*-value 9.42E-03), and the two most downregulated genes bar *dsdC*, *c3818* and *c5343* had fold changes of -4.6 (*p*-value 0.04) and -2.76 (*p*-value 0.03). When the two D-serine data sets were overlapped, there were 56 genes that were commonly differentially expressed, although only 6 of these were transcriptionally differentially expressed in the same direction: *metL*, *bfr*, *yfiD*, *dsdC*, *c3149*, and *c2566*.

There was also a difference in the number of genes differentially expressed between the CE10 "WT$^{D\text{-}ser}$ vs. ΔΔ*dsdC1/2* + pDsdA1/2$^{D\text{-}ser}$" data set, in comparison to "WT$^{D\text{-}ser}$ vs. Δ*dsdC*$^{D\text{-}ser}$"; the number of DEGs increased from 436 to 552, with 38 of these DEGs having fold changes of >3 and <-3. The two most significant upregulated genes bar *dsdA1* and *dsdA2*, *gadE* and *yfiD* had fold changes of 16.68 (*p*-value 1.45E-04) and 11.59 (*p*-value 8.38E-90). Notably, in this pairwise comparison *dsdC2* was downregulated -210.52-fold (*p*-value 1.61E-07), whereas *dsdC1* was downregulated -36.41 (*p*-value 2.32E-03). The two most downregulated genes bar *dsdC1* and *dsdC2*, *CE10_4853* and *mdtJ,* had fold changes of -161.47 (*p*-value 2.44E-13) and -11.2 (*p*-value 7.77E-07). These results indicated that addition of pDsdA to ΔΔ*dsdC1/2*$^{D\text{-}ser}$ caused more differential gene expression, suggesting that overexpression of pDsdA put the cell under undue stress. When the two CE10 D-serine data sets were overlapped, this revealed 124 common DEGs, although only 30 of these were transcriptionally differentially expressed in the same direction (Table 3-5).

	Condition 2 CE10 WT$^{D\text{-}ser}$ vs. ΔΔ*dsdC1/2*$^{D\text{-}ser}$		Condition 3 CE10 WT$^{D\text{-}ser}$ vs. ΔΔ*dsdC1/2* + pDsdA1/2$^{D\text{-}ser}$	
	Fold change	**FDR *p*-value**	**Fold change**	**FDR *p*-value**
kpsT	3.03	2.08E-04	1.66	1.04E-04
neuA	4.32	1.99E-06	2.01	1.11E-08
neuB	6.93	9.96E-08	1.99	6.19E-09
neuE	4.27	6.44E-04	3.1	3.41E-11
neuC	6.57	1.99E-06	2.31	2.56E-11
neuS	6.89	1.30E-04	3.98	2.71E-16
neuO	-4.57	8.56E-10	-8.15	5.07E-71
papA2	2.46	6.43E-04	1.57	2.50E-03
metQ	1.59	1.65E-04	2.5	5.03E-07
wbbD	10.02	5.74E-04	1.98	2.00E-02
wbbA	6.17	1.02E-03	1.98	2.00E-02
rfaY	2.33	1.00E-02	1.73	8.36E-03
wbbC	7.38	8.82E-05	2.12	1.35E-03
wzx	7.6	2.05E-05	2.29	3.87E-04
waaW	2.04	1.00E-02	1.88	1.67E-04
waaT	2.31	1.30E-03	2.06	1.13E-05
vioA	3.71	5.23E-04	2.25	1.48E-07
waaV	2.12	3.00E-02	4.25	2.40E-10
gudD	4.29	5.68E-06	1.68	2.00E-02
hdhA	2.39	3.06E-03	1.64	3.07E-03
metL	3.05	2.70E-09	1.55	6.10E-04
metB	5.32	3.64E-19	1.7	3.93E-05
CE10_1649	-2.9	1.77E-05	-3.17	3.31E-14
repB	1.95	4.00E-02	1.59	2.19E-03
gnsB	7.02	7.90E-06	2.69	6.58E-11
dsdC2	-74.42	1.15E-05	-210.52	1.61E-17
yjiY	21.21	2.55E-38	1.82	1.00E-02
dppA	1.65	3.00E-02	1.53	6.36E-04
CE10_1029	2.09	4.51E-03	1.51	8.11E-03
CE10_3484	7.12	1.67E-07	2.54	6.15E-06

Table 3-5 Commonality of DEGs between CE10 condition 2 and condition 3 datasets. Fold change and FDR *p*-values for the 30 common genes that overlapped between the CE10 "WT$^{D\text{-}ser}$ vs. ΔΔ*dsdC1/2*$^{D\text{-}ser}$" and "WT$^{D\text{-}ser}$ vs. ΔΔ*dsdC1/2* + pDsdA1/2$^{D\text{-}ser}$" pairwise comparisons. Genes that had an increase in fold change values are shown in green, and genes that had a decrease in red.

GO analysis was used to assess the functionality of the DEGs for the pairwise comparisons. The combined datasets of condition 2 and condition 3 ("WT$^{D\text{-}ser}$ vs. Δ*dsdC*$^{D\text{-}ser}$" and "WT$^{D\text{-}ser}$ vs. Δ*dsdC* + pDsdA$^{D\text{-}ser}$") were used. Figure 3-24 shows the up- and down-regulated genes for the pairwise comparisons in their functional groups as determined by UniProtKB.

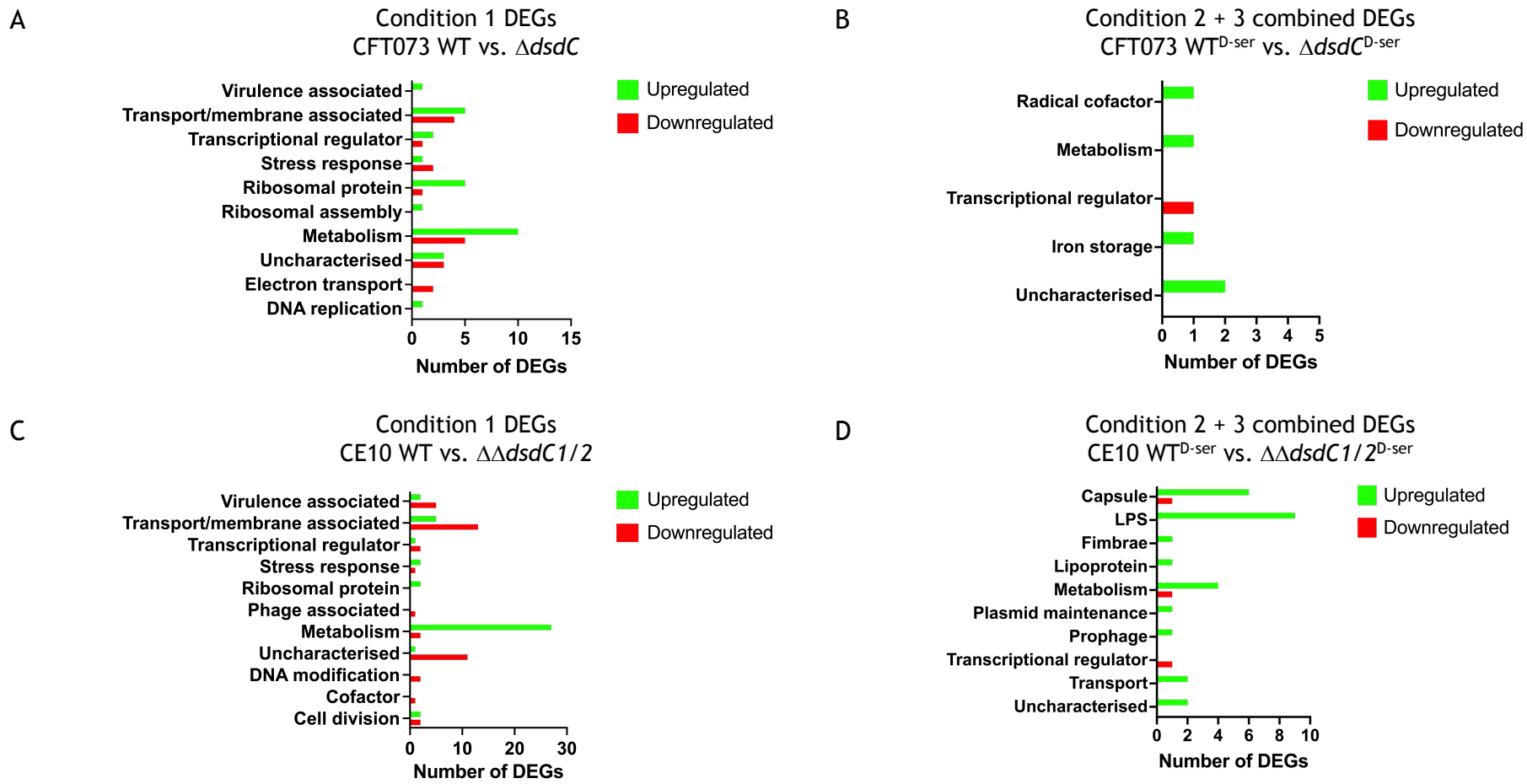


Figure 3-24 Functional characteristics of the DEGs using the CFT073 and CE10 transcriptomic data. Genes that were differentially expressed (*p*-value ≤0.05) were characterised using GO analysis from UniProtKB into broad functional groups; (A) CFT073 WT vs. Δ*dsdC*; (B) combined gene list from condition 2 and 3 (CFT073 $WT^{D\text{-}ser}$ vs. Δ$dsdC^{D\text{-}ser}$); (C) CE10 WT vs. ΔΔ*dsdC1/2*; (D) combined gene list from condition 2 and 3 (CE10 $WT^{D\text{-}ser}$ vs. ΔΔ$dsdC1/2^{D\text{-}ser}$)

Figure 3-24 shows in the absence of D-serine, functional group patterns largely indicated that metabolism was one of the most affected functional groups. Indeed, in the CFT073 data set 15 of the 47 DEGs were involved in metabolism processes. Further functional analysis revealed DEGs involved in DNA replication (1/47), electron transport (2/47), ribosomal assembly (1), ribosomal proteins (6/47), stress response (3/47), transcriptional regulation (3/47), transport and membrane associated (9/47), virulence associated (1/47), and uncharacterised (6/47). This was also reflected in CE10, were 29 of the 82 DEGs were metabolism related. Similarly to CFT073, many DEGS were involved in transport or were membrane associated (18/82). Functions also included cell division (4/82), cofactors (1/82), DNA modification (2/82), phage associated (1/82), ribosomal proteins (2/82), stress response (3/82), transcriptional regulators (3/82), virulence associated (7/82), and uncharacterised (12/82). These results indicated that in the absence of D-serine, DsdC regulates core processes of the cell. Indeed, only 1 DEG was associated with virulence in the CFT073 data set when D-serine was absent, *fimG* (1.88-fold change; *p*-value 1.5E-03), a protein involved in cell adhesion.

Overlap of the CFT073 D-serine transcriptomic data sets had revealed only 6 DEGs that were transcriptionally differentially expressed in the same direction. GO analysis of these 6 DEGs revealed no functional group patterns: *metL* (lysine biosynthesis), *bfr* (bacterioferritin), *yfiD* (putative radical cofactor), *dsdC* (transcriptional regulator), *c3149* (hypothetical), and *c2566* (hypothetical) (Fig. 3-24B). Strikingly, GO analysis of the DEGs overlapped in the CE10 D-serine data sets revealed a link between D-serine and virulence (Fig. 3-24D). Indeed, there were three functional groups that were associated with NMEC virulence: capsule (7/30), LPS (9/30), and fimbriae (1/30). Further functions also included genes involved in lipoprotein (1/30), metabolism (5/30), plasmid maintenance (1/30), prophage (1/30), transcriptional regulator (1/30), transport (2/30) and uncharacterised (2/30). These results indicated that DsdC in the presence of D-serine is involved in the regulation of capsule and LPS, two important virulence determinants in NMEC.

Global transcriptome profiling using RNA-Seq had revealed the novel regulon of DsdC, in the presence and absence of D-serine. Intriguingly, genes involved in core processes like cell division and ribosomal proteins were differentially

expressed when D-serine was absent. However, in the presence of D-serine in CE10, the regulon became virulence associated with differential expression of capsule and LPS biosynthesis genes. These results may indicate that D-serine is being used by the bacteria as a signal involved in virulence associated mechanisms. To establish if the DEGs were caused by direct DsdC regulation, or if this was due to indirect effects such as DsdC involved in the regulation of another TF or the toxic effects of D-serine, the ChIP-Seq binding data and the RNA-Seq transcriptomic data was compared.

3.5.2 DsdC directly regulates a small, but virulence associated, regulon in NMEC strain CE10

To understand the direct role of DsdC on gene expression in CFT073 and CE10, the binding and transcriptomic datasets were compared. The comparisons of the data are presented in Figure 3-25 and the common genes list is presented in Supplementary Table S8.

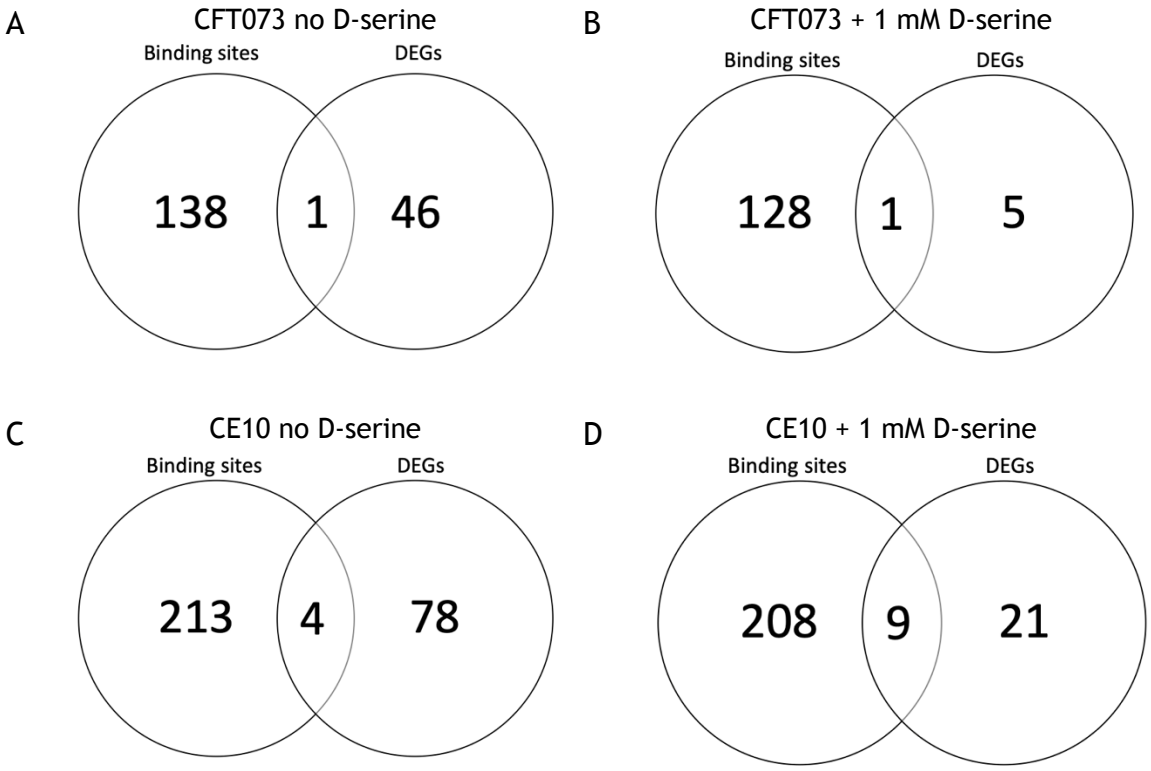


Figure 3-25 Direct regulation of DsdC in CFT073 and CE10. Venn diagram showing commonality between the binding sites in the ChIP-Seq data compared to the differential gene expression from the transcriptomic data for DsdC. (A)

CFT073 with no D-serine, (B) CFT073 + 1 mM D-serine, (C) CE10 with no D-ser and (D) CE10 + 1 mM D-serine.

Figure 3-25 revealed there was little overlap between the DsdC binding and transcriptomic data. Indeed, for CFT073 DsdC, in both the presence and absence of D-serine, the only overlap between the datasets was *dsdC* (Fig. 3-25(A-B)). These results therefore indicated that DsdC plays no direct role in regulation in CFT073, bar D-serine metabolism. For CE10 (Fig. 3-25C), when D-serine was absent there were only 4 genes that appeared to be directly regulated by DsdC: *dsdC2*, *ypdI* (putative lipoprotein), *gadB* (acid tolerance) and *neuO* (capsule). When D-serine was present this increased to 9, and solely included genes involved in capsule production and LPS biosynthesis bar *dsdC2*: *neuA, neuB, neuO, rfaY, wbbC, wzx, waaT,* and *waaV*. These results indicated a direct role of DsdC in CE10 gene regulation in both the presence and absence of D-serine. Strikingly, in the presence of D-serine, a direct role of DsdC in the regulation of capsule and LPS expression emerged, two key virulence mechanisms in NMEC. These results could therefore indicate that DsdC has been tailored in CE10 to regulate a strain-specific phenotype.

3.6 Discussion

3.6.1 DsdC is absolutely required for D-serine metabolism but has no effect on growth in general

As has been reported extensively elsewhere, DsdC is absolutely required for D-serine metabolism (Nørregaard-Madsen *et al.*, 1995). The data presented here concurs, showing that D-serine concentrations relative to host levels inhibited the growth of the Δ*dsdC* mutants completely, with growth only restored with complementation of pDsdC (Fig. 3-3, 3-4, and 3-5). Further, deletion of the regulator did not appear to affect growth, with no defects reported when Δ*dsdC* was grown in minimal media. This correlates with data reported in *P. mirabilis,* where deletion of the operon did not affect growth where other carbon and nitrogen sources were available (Brauer *et al.*, 2019).

Gene duplication is present in many bacterial genomes. Indeed, 94 genes in CFT073 are duplicated, with some of these duplications revealed to be

pathotype specific (Bernabeu *et al.*, 2019). 30 out of 41 K1 strains tested were reported to carry two copies of the *dsdCXA* operon, with both copies functional in D-serine metabolism (Moritz & Welch, 2006). CE10 also carries two unlinked copies of the *dsdCXA* operon. To establish functional redundancy of D-serine metabolism between the two copies, single Δ*dsdC1* and Δ*dsdC2* mutations were made. Figure 3-3 and 3-4 revealed both DsdC1 and DsdC2 were functionally redundant in regard to D-serine metabolism. Further, when the native levels of expression of both DsdC1 and DsdC2 were analysed, there appeared to be no difference in protein expression (Fig. 3-8). Gene duplications are often associated with adaption of cells to changing environments (Bernabeu *et al.*, 2019). As stated in 3.2.1 there are small changes in the amino acid structure between CE10 DsdC1 and DsdC2, with most changes occurring in a small cluster in the predicted substrate binding domain. Speculatively, these differences could cause changes in the tertiary structure of the protein, potentially altering the specificity of the binding pocket and allowing a different substrate to bind. DsdC1 and DsdC2 could therefore be recognising different co-inducers in CE10. Although this is an interesting hypothesis, determining the specific co-inducers for DsdC was not carried out in this work. However, as the ChIP-Seq global binding data revealed, under the conditions that were used, there were no clear differences in differential binding sites between DsdC1 and DsdC2, suggesting therefore that they are likely to play a redundant role in CE10.

3.6.2 DsdC is a global transcription factor that binds to multiple regions on the genome

Although Anfora *et al.*, 2007, speculated a larger role in gene regulation for DsdC in ExPEC strains; this is to our knowledge, the first report that DsdC plays a role in *E. coli* gene regulation, beyond D-serine metabolism. Analysis of the ChIP-Seq binding data revealed that DsdC bound to multiple regions of the genome, with CE10$^{DsdC1\text{-}FLAG}$ in the presence of D-serine, binding to 217 regions (Table 3-3; Supplementary Table S1-6). As Maddocks and Oyston (2008) reviewed, LTTRs can be local or global regulators, and indeed previous studies on LTTRs have revealed a range in the number of binding sites. YhaJ, a LTTR in *E. coli* had shown binding to 7 peaks in CFT073, whereas MetR, a LTTR in *Streptococcus pneumoniae* had shown binding to 52 loci of the genome, and OxyR, a LTTR in *Pseudomonas aeruginosa* showed binding to 122 regions

(Connolly *et al.*, 2019; Wei *et al.*, 2012; Zhang *et al.*, 2021). Strikingly, NAPs such as Fis and H-NS can bind up to 1,464 and 458 regions of the *E. coli* genome respectively (Kahramanoglou *et al.*, 2011). The number of ChIP peaks generated from this experiment was more akin to a global LTTR and indicated that DsdC is not just binding to the *dsdCX* region but is binding throughout the genome.

The canonical TF model, based on early work by Jacob and Monad on the lac operon, is that TFs bind within the TSS at the 5' region of a gene, usually in intergenic non-coding regions. Many of the DsdC ChIP-Seq peaks fitted this model, although ~25% of the ChIP-Seq peaks obtained were binding in intragenic regions. Singh *et al.*, (2014) reported that in the *E. coli exhCABD* operon, there were multiple intragenic promoters. Further, mapping of the NikR regulon using a combination of ChIP-Seq and RNA-Seq in *Helicobacter pylori* revealed intragenic binding, with 41 out of 72 peaks positioned within annotated genes (Vannini *et al.*, 2017). This phenomenon does not appear to be limited to bacteria either. Research in archaeal operons and coding sequences revealed the presence of intragenic promoters, with 11 TFs reported to bind inside operons and annotated genes (Koide *et al.*, 2009). Intragenic binding therefore appears to be widespread.

Intragenic binding of TFs could have arisen due to a limitation on space within the chromosome or due to the TF previously regulating a gene that has since been lost due to chromosomal rearrangement (Galagan *et al.*, 2012). Alternatively, binding in these intragenic regions could be to prevent transcriptional elongation or hinder processing. TF binding site (TFBS) positions in relation to the promoter have been suggested to be an important factor in determining regulatory function, with TF repressors more likely to bind downstream of the promoter than TF activators (Babu & Teichmann, 2003b). One study showed that a third of repressor binding sites occurred after the TSS (Babu & Teichmann, 2003b). This could suggest that DsdC may be binding in intragenic regions of the gene to hinder transcription. Although intergenic peaks were considered the norm for transcriptional regulation, an unexpected observation of ChIP-Seq data has been that many TFBS occur within the ORF (Wade, 2015). The results presented here correlate with that finding, with DsdC showing binding in both intergenic and intragenic sites.

Further genomic analysis of the ChIP-Seq hits revealed that DsdC bound mainly in the TSS of the gene, in the '0 bp bin'. This is typical of a LTTR which often binds at -35 to +20 bp sites (Maddocks & Oyston, 2008). However, there was a range in distance of DsdC binding sites, with 7 TFBS being further than 150 bp upstream from the TSS in CFT073$^{DsdC\text{-}FLAG}$. Although unusual, this has been reported in prokaryotes before, with several TFs binding >150 bp upstream of the TSS (reviewed in Galagan, Lyubetskaya & Gomes, 2012). Indeed, AraC binds 280 bp upstream to the P_{BAD} TSS and mediates gene expression by DNA looping (Dunn *et al.*, 1984). A further study revealed that VjbR, a LuxR-type transcriptional regulator, bound to the genome >300 bp upstream of the TSS; although, there was preferential binding between 20 and 140 bp upstream (Kleinman *et al.*, 2017). The data presented here therefore seems consistent with previous reports of TFBS.

3.6.3 DsdC binds in a strain-specific manner in two ExPEC pathotypes

As the DsdC ChIP-Seq data had revealed multiple binding sites on the genome, functional characteristic analysis of the putatively bound genes was carried out. This analysis revealed that DsdC bound to genes involved in multiple different processes ranging from metabolism to electron transport and even virulence, suggesting that DsdC plays a global role in two ExPEC strains, and does not simply regulate one functional group. This is also the case for the LTTR OxyR. A ChIP-exo experiment in *E. coli* MG1665 revealed, under oxidative stress conditions, OxyR bound to genes involved in detoxification and DNA damage repair, as well as amino acid biosynthesis, cell wall synthesis and metal ion transport (Seo *et al.*, 2015). Furthermore, a ChIP-chip experiment of the LTTR, LeuO, revealed binding to 178 genes on the genome of *Salmonella enterica* serovar Typhimurium, with gene functions ranging from housekeeping to virulence-associated (Dillon *et al.*, 2012).

GO functional analysis upon the CE10 DsdC ChIP-Seq hits revealed similar binding patterns to CFT073; however, there were also differences suggesting that DsdC has been tailored for strain-specific regulation. Indeed, in CFT073 there was binding to *tssB* and *yheE*. *tssB* and *yheE* (*gspC* in K12) are genes involved in the T6SS and a putative T2SS respectively (Korotkov *et al.*, 2011; Navarro-Garcia *et*

al., 2019). These are both mechanisms utilised by bacteria to secrete virulence effectors into the host or close neighbouring bacteria, with one study revealing, in UPEC, the T2SS was important for persistent infection in the upper urinary tract (Kulkarni *et al.*, 2009). Although NMEC reportedly has an orthologue of the T6SS (Navarro-Garcia *et al.*, 2019), there were no homologues of *tssB* nor *yheE* found in CE10. Furthermore in CFT073, DsdC was shown to bind to *clbR*, the colibactin master regulator (Wallenstein *et al.*, 2020). Colibactin is a bacterial genotoxin found on the *pks* island and has been linked with colorectal cancer (CRC). Although many B2 *E. coli* strains carry the *pks* island (Faïs *et al.*, 2018), CE10, which is in phylogroup D, does not carry it.

In CE10 it was shown that DsdC uniquely bound to genes involved in the K1 capsule and the putative T3SS, ETT2. The ETT2 in the past has been termed 'cryptic' due to lack of understanding of what role it played. However, it has been suggested to play an important role in bloodstream isolates of *E. coli*, with mutations in CE10 of the ETT2 leading to defects in invasion and intracellular survival in HBMECs (Yao *et al.*, 2009). Recently Fox *et al.*, (2020), analysed bloodstream *E. coli* isolates and found that strains from the sequence type (ST) 69 lineage contained an intact ETT2, indicating its importance in virulence. As CFT073 does not carry these genes, these results taken together, indicate that DsdC has been tailored in ExPEC strains, mediating different regulatory pathways that are important to the individual strains requirement.

The central dogma for many years has been that TFs regulate the same set of genes in different strains of the same species, indeed even in different bacteria. Perez and Groisman (2009a) revealed that the homologous TF PhoP was shown to regulate the same 'ancestral' set of genes in two different bacterial species, *S.* Typhimurium and *Yersinia pestis*. However, recent studies have started showing the potential repurposing and tailoring of TFs for strain-specific regulation. Recent work from the Roe group has shown that the conserved TF YhaJ directly regulates strain-specific virulence mechanisms, regulating expression of T3SS components and effectors in EHEC, whilst in UPEC regulating T1F expression (Connolly *et al.*, 2019). Indeed, even homologous TFs have recently been shown to regulate different pathways. *Xanthomonas campestris* pv *campestris* carries two sigma factor 54 proteins, RpoN1 and RpoN2. Recent data indicates that these two homologues have distinct roles in pathogenesis in *X. campestris*, with

RpoN2 regulating flagellation, and RpoN1 potentially regulating fatty acid synthesis (Li *et al.*, 2020). Although this is an important observation, it should also be noted that these two proteins share only 56% of residues.

The results presented here indicate that there is strain-specific binding in two ExPEC pathotypes, by the conserved TF DsdC. The relative plasticity, but finite size, of the *E. coli* genome has led to several distinct *E. coli* pathotypes emerging; indeed, CFT073 and CE10 both utilise distinct virulence mechanisms whilst infecting the host. As they both are able to use D-serine as a carbon source, DsdC likely plays a role in the success of these two ExPEC strains. Further, it is likely, due to the need of newly acquired genes to be integrated into 'ancestral' networks (Perez & Groisman, 2009b), DsdC has been tailored to bind in a strain-specific manner.

3.6.4 DsdC directly regulates a small subset of virulence-associated genes in NMEC strain CE10

As direct binding does not always equate to a transcriptional response, a comparison between the global binding data and the transcriptomics experiment was made. Intriguingly, there was little direct regulation of genes observed when comparing the ChIP-Seq and RNA-Seq data. DsdC is therefore binding to the genome but not causing a direct regulatory effect upon many genes. This was initially surprising, although several other studies have reported results of TF binding but no transcriptional responses being observed. ChIP-chip analysis of the TF RutR in *E. coli* strain BW2511 revealed 14 out of the 20 peaks observed were intragenic, and appeared to play little to no role on transcription (Shimada *et al.*, 2008). In *H. pylori*, a combined ChIP-Seq and RNA-Seq analysis of NikR revealed the presence of 'transcriptionally orphaned' NikR binding sites, with no apparent role in transcription (Vannini *et al.*, 2017). Further, in *B. abortus*, ChIP-Seq analysis combined with RNA-Seq, revealed binding of VjbR to 235 regions of DNA, with only 37 of those correlating with differentially expressed genes (Kleinman *et al.*, 2017).

As the *E. coli* genome is ever-evolving, it has been suggested that these non-regulatory bindings could be evolutionary relics, with the genes the TF used to regulate since 'lost' from the genome (Shimada *et al.*, 2008). Moreover, it has

also been suggested that there could be binding on the genome at non-regulatory sites so that the cell can maintain an optimum amount of TF present (MacQuarrie *et al.*, 2011). Indeed, it has been described for the *lac* repressor that up to 98% is bound to sites other than the *lac* operator (Lin & Riggs, 1975). This hypothesis was also proposed by Vannini *et al.*, (2017), theorising that NikR bound to the genome at dedicated 'parking bays', in order to maximise the concentration of the regulator in certain regions. Furthermore, one study that made a genome-wide reconstruction of the OxyR/SoxRS regulon in *E. coli* MG1655, suggested that co-regulation could play a role in not seeing a transcriptional effects; with co-regulating TFs taking over the regulatory role when the relevant TF is missing (Seo *et al.*, 2015).

There were also limitations within the experiments presented in this thesis. Due to the nature of the project, the two global experiments were carried out separately, with the RNA-Seq experiment performed prior by Dr Nicky O'Boyle. Although every effort was made to limit the differences in experimental setup, small discrepancies between the procedures may have arisen, potentially contributing to the differences between the binding and transcriptomic data. To minimise variation for RNA-Seq, Gao *et al.*, (2021) performed all biological replicate experiments on the same day. Further, the same samples were then used for observing bacterial motility by microscopy and video analysis, enabling a direct comparison of the two experiments (Gao *et al.*, 2021). The technique of using the same samples for combination experiments, should potentially be used for future comparison experiments, thereby limiting any possible variation observed.

Further, there is the possibility that the RNA-Seq samples were taken at a time point where the transcriptional effects of DsdC, after induction of D-serine, were no longer being observed. *Vibrio coralliilyticus* has been reported to differentially regulate 2,705 genes, 10 minutes after exposure to an inducer (Gao *et al.*, 2021). After 60 minutes, the number of DEGs dropped to 2,235 genes (Gao *et al.*, 2021). Strikingly, only 30% of DEGs were shared between the 10- and 60-minute data sets, indicating that sampling time is a relevant factor when making observations of transcriptional responses. Another recent study that observed the transcriptional effects of the regulator ppGpp, identified a transcriptional response in 757 genes 5 minutes after induction (Sanchez-

Vazquez *et al.*, 2019). This data therefore begs the question of whether the RNA-Seq sampling time point (120 minutes post-induction) was potentially taken too late, and had the time point been taken earlier after induction, would more overlap have been seen between the two global sequencing data sets? Future studies could answer this by sampling and performing RNA-Seq or RT-qPCR studies over a gradient of time post-induction.

The large number of non-regulatory binding sites, compared to the relatively low number of regulatory binding sites, could also be down to the nature of the ChIP-Seq experiment. ChIP-Seq data is known to produce false positives (Pickrell *et al.*, 2011); indeed, one study using *Drosophila melanogaster* embryos revealed 3,000 non-specific ChIP-Seq peaks, that they described as 'Phantom Peaks' (Jain *et al.*, 2015). Marx (2019) discussed the occurrence of phantom ChIP-Seq peaks which can arise from non-specific antibody binding at promoter elements. Further, some regions of the genome are 'hyper-ChIPable', leading to false positive data outputs (Marx, 2019). However, as a non-tagged control was used in the DsdC analysis, the number of false positives obtained in this data set should be limited.

In CFT073, in both the presence and absence of D-serine, the only gene common between the binding and transcriptomic data sets was *dsdC*. Conversely in CE10, in the absence of D-serine, DsdC bound and showed differential expression of 4 genes. GO analysis of these genes revealed there was no functional group consensus that DsdC is binding to. Strikingly, in the presence of D-serine for CE10, all of the genes that overlapped in the global binding and transcriptomic data sets were involved in virulence, bar *dsdC*2. Indeed, the 8 genes were all involved in LPS and K1 capsular biosynthesis, two major virulence determinants of NMEC. This data indicated that DsdC directly regulates a small, but virulence-associated regulon in CE10.

3.6.5 The indirect regulon of DsdC

As discussed above, at the time points sampled, there was little overlap between the binding and transcriptomic data sets, indicating that DsdC indirectly regulates multiple processes within the cell. In the absence of D-serine DsdC was found to, indirectly, differentially regulate 46 and 78 genes in

CFT073 and CE10 respectively. Upon exposure to D-serine, 1,055 and 411 genes were, indirectly, differentially expressed in CFT073 and CE10 respectively. This data indicated that DsdC was more transcriptionally active in the presence of its suspected co-inducer, D-serine, which is comparable with other LTTRs (Maddocks & Oyston, 2008). However, the increased number of DEGs could also be due to the toxic side effects of D-serine accumulation in the Δ*dsdC* background. DsdC is required for expression of *dsdA*, the D-serine deaminase, which breaks down D-serine into pyruvate and ammonia (Nørregaard-Madsen *et al.*, 1995). In the absence of DsdA, D-serine becomes toxic to the cell by inhibiting the L-serine and pantothenate pathway (Cosloy & McFall, 1973; Maas & Davis, 1950). By including the pDsdA plasmid in the transcriptomics experiment it was hoped to alleviate the problem of D-serine accumulation and corresponding toxicity.

Indeed, in the CFT073 data set the number of DEGs dropped dramatically from 1,067 to 174. However, upon comparing the two data sets it was revealed that only 6 were transcriptionally differentially expressed in the same direction. This could be an accurate representation of the DsdC regulon, although, it was also considered the possibility that overexpression of pDsdA was exerting a toxic effect within the cell. Conversely, in CE10 the number of DEGs increased from 436 to 552 when pDsdA was overexpressed. When the two data sets were compared only 30 genes were transcriptionally differentially expressed in the same direction. Therefore, it is possible that the overexpression of pDsdA is causing a metabolic burden upon the cell, thus affecting the transcriptomic response (Bolognesi & Lehner, 2018). As the primary interest of this work was in elucidating the direct role of DsdC, the remainder of the work presented in this thesis will focus upon the genes that were present in both the global binding data and the transcriptomics data sets.

3.6.6 DsdC as a canonical transcription factor and a nucleoid associated protein

Intriguingly, two distinct patterns emerged from the ChIP-Seq binding data: peaks that were canonical in shape, and peaks that were broader and covered larger sections of the genome. The canonical binding profile of a TF is that it binds as a ‘narrow’ peak in the TSS of the gene it is regulating, whereas

‘broader’ peaks are usually associated with histones (Starmer & Magnuson, 2016). Figure 3-13 highlighted the differences in these binding profiles with the narrow binding peak spanning a few hundred base pairs whereas the broader peak spanned much larger regions, in the case of the LPS biosynthesis region >6,000 bp. Although most prokaryotes do not contain histones, NAPs have been described previously as being ‘histone-like’ in manner (Dillon & Dorman, 2010). Indeed, the NAP H-NS often binds to longer regions of DNA; a study in *E. coli* showed binding up to 1,759 bp, whereas in *S.* Typhimurium the average binding region was 1,500 bp, however it had been noted to increase to 10,750 bp (Dillon *et al.*, 2010; Kahramanoglou *et al.*, 2011). As it appears for DsdC, H-NS also has the ability of binding in longer and shorter regions. Kahramanoglou *et al.*, (2011) found that the shorter regions tended to behave as canonical TFs whereas the longer regions enfolded large regions of the genome, likely acting as a transcriptional silencer.

These results could indicate that DsdC is acting in both a classical TF manner and a NAP. Indeed, the lines between the two have often been blurred (Dorman *et al.*, 2020), and a recent paper redefining fundamental concepts in transcriptional initiation has suggested that as NAPs often have functions that overlap TFs, they should no longer be defined as two separate classes (Mejía-Almonte *et al.*, 2020). Therefore, it is entirely plausible that DsdC is acting both as a TF and as a NAP. This would also correlate to the ChIP-Seq binding data that indicates that ~25% of the peaks are binding in intragenic regions. NAPs such as H-NS are known to bind intragenic regions, silencing transcription (Singh *et al.*, 2014). It could also explain why many of the ChIP-Seq binding hits are not associated with direct regulatory functions. Hypothetically DsdC is acting more like a NAP and is involved in chromosomal organisation, and is therefore binding along the genome in order to potentially fine tune bending, bridging or wrapping of DNA (Dame, 2005). This hypothesis could also explain why a classical T-N_{11}-A motif was not obtained when the sequence data was analysed on MEME. NAPs are known to bind to the genome with low sequence specificity (Wade, 2015), and it was not until canonical TF binding peaks were used that the MEME software generated a statistically significant motif. Indeed, LTTRs are known to rely on DNA shape rather than on nucleotide sequence in recognising their DNA targets, a trait that is shared by NAPs (Dorman *et al.*, 2020). Given this data,

this could indicate that DsdC potentially acts as both a canonical TF and as global NAP.

3.6.7 Conclusion

To our knowledge, the work presented here is the first evidence that DsdC plays a role in *E. coli* beyond regulating the D-serine metabolism locus. Using global binding and transcriptomic data, it has been revealed that DsdC binds globally across the CFT073 and CE10 genome. Further, it has been shown that although there were many overlaps, DsdC bound in a strain-specific manner in two ExPEC strains, binding to colibactin synthesis genes in CFT073 and ETT2 effector genes in CE10. Moreover, using transcriptomics it was shown that DsdC differentially regulates multiple genes outside the *dsdCXA* locus. Strikingly, overlap of the transcriptomic and binding data in CE10 revealed a small subset of genes that were common to both sets, all involved in LPS and K1 capsular biosynthesis. These results together indicate, that over time, DsdC has been tailored for strain-specific regulation of virulence factors.

Chapter 4 Validation of the global sequencing data and purification of the LysR-type transcriptional regulator, DsdC

4.1 Introduction

Regulatory control of genes in response to environmental cues is central to bacterial survival (Browning & Busby, 2016). Indeed, the rapid recognition of environmental signals and subsequent change in gene expression allows bacteria to survive in fluctuating environmental conditions (Shimizu, 2013). Control of bacterial gene expression is a multifaceted process, with multiple regulatory mechanisms that can influence both transcription and translation (Browning & Busby, 2004; Starosta *et al.*, 2014). Bacteria that are pathogenic, or that can survive in multiple niches within the host, often carry larger numbers of regulatory elements, including TFs, reflecting their adaptability to diverse host environments (Bervoets & Charlier, 2019). UPEC and NMEC are two versatile ExPEC pathotypes that can cause serious disease in humans (Kaper *et al.*, 2004). ExPEC strains predominantly reside asymptomatically within the gut, causing disease upon egression (Johnson & Russo, 2002). Within the urinary tract, UPEC strains often survive in a nutritionally depleted environment, containing mostly amino acids and small peptides, whereas NMEC strains can translocate from the mucosal epithelium to the bloodstream and traverse the BBB into the CNS (Alteri *et al.*, 2009; Silver & Vimr, 1990). Although UPEC and NMEC can reside in distinct niches of the human body, they both utilise D-serine as a carbon source. In *E. coli*, DsdC is a TF that regulates the D-serine metabolism locus.

In Chapter 3, results from global binding and transcriptomic experiments demonstrated that DsdC bound to a large number of genetic regions in prototypic UPEC and NMEC strains and caused differential gene expression upon exposure to D-serine. This chapter aims to validate the RNA-Seq data presented in Chapter 3 using RT-qPCR analysis, demonstrating that DsdC affects expression of genes involved in capsular and LPS biosynthesis. Moreover, to explore the ChIP-Seq binding results presented in Chapter 3 and to demonstrate that DsdC directly binds to regions of the CE10 and CFT073 genome, DsdC (CFT073), DsdC1 (CE10) and DsdC2 (CE10) were purified. As LTTRs are generally known to be insoluble proteins, a maltose binding protein (MBP) tag was used to enhance the solubility of the DsdC fusion protein (Sachdev & Chirgwin, 2000). Furthermore, DsdC1 (CE10) was also tagged with a 6xHis tag. The purified DsdC was then used in electrophoretic mobility shift assays (EMSA) to confirm binding of DsdC to its own promoter and also to an LPS glycosyltransferase gene, *waaV*. Finally, the

functionality of a TF often relies on the recognition of specific DNA sequences at a promoter (reviewed in Browning, Butala & Busby, 2019). The binding motif of DsdC is currently unknown and therefore this chapter aimed, using DNase I footprinting and purified DsdC, to elucidate the binding motif.

4.2 Validating the RNA-Seq transcriptomic data using RT-qPCR

As discussed in Chapters 1 and 3, global transcriptome profiling by RNA-Seq can identify the regulon of a TF, measuring differences in RNA levels between WT and isogenic mutants. RT-qPCR is often used as validation for RNA-Seq data and is a sensitive method that measures gene quantification (Fang & Cui, 2011). The RNA-Seq data presented in Chapter 3 revealed a number of genes that were differentially expressed in the CE10 ΔΔ*dsdC1/2* background upon exposure to D-serine.

Several of these genes were related to LPS biosynthesis and K1 capsular production. Validation of the RNA-Seq data was therefore carried out on genes of biological interest, including *neuB, waaV, kpsF* and *neuO*. RNA was extracted from CE10 WT and ΔΔ*dsdC1/2* isogenic mutants. The samples were cultured in M9 minimal media for 3 hours and spiked with 1 mM D-serine and left to incubate for a further 2 hours. RT-qPCR analysis on three biological replicates confirmed and validated the RNA-Seq results. Upon exposure to D-serine, gene expression of *neuB* (1.56-fold; *p*-value 0.000038), *waaV* (1.36-fold; *p*-value 0.0091), and *kpsF* (1.91-fold; *p*-value 0.0042) were significantly increased in the ΔΔ*dsdC1/2* mutant compared to CE10 WT, whereas gene expression of *neuO* was decreased (-2.36-fold; *p*-value 0.026) (Table 4-1; Fig. 4-1). These results validated the RNA-Seq data, indicating that DsdC affects the regulation of genes involved in CE10 LPS and capsular biosynthesis.

Gene	RNA-Seq		RT-qPCR	
	Fold change	*p*-value	Fold change	*p*-value
neuB	1.99	6.19E-09	1.56	3.8E-05
waaV	2.12	0.03	1.36	9.1E-03
kpsF	1.71	9.07E-03	1.91	4.2E-03
neuO	-4.57	8.56E-10	-2.36	2.6E-02

Table 4-1 Comparisons of the RNA-Seq and RT-qPCR data. A direct comparison of the fold changes and respective *p*-values from the RNA-Seq and RT-qPCR data for several biologically relevant genes; *neuB, waaV, kpsF* and *neuO*, for which fold change was measured in the CE10 Δ*dsdC1/2* mutant relative to the CE10 WT, in the presence of D-serine.

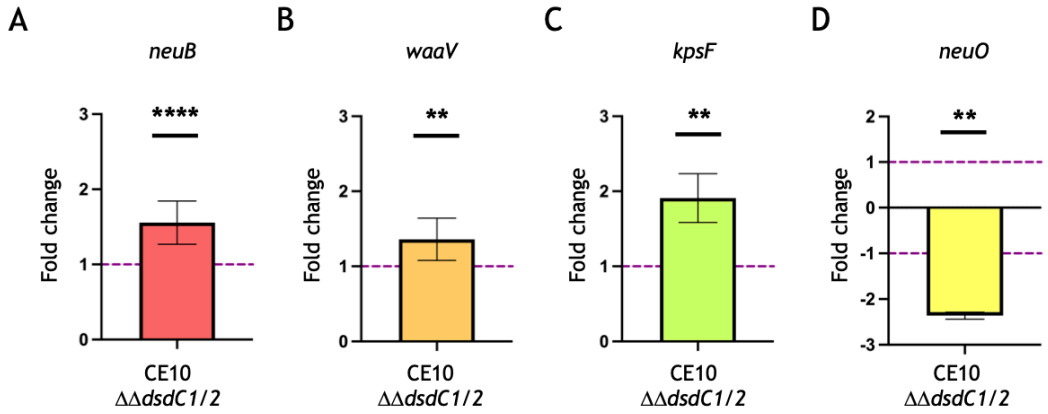


Figure 4-1 RT-qPCR analysis of capsular and LPS biosynthesis genes. RT-qPCR validation of the RNA-Seq data showed increased fold changes for (A) *neuB*, (B) *waaV*, and (C) *kpsF*, in the CE10 Δ*dsdC1/2* mutant relative to the CE10 WT, in the presence of D-serine. (D) shows the decreased fold change in *neuO* in the CE10 Δ*dsdC1/2* mutant relative to the CE10 WT, in the presence of D-serine. The purple dashed line indicates baseline expression of each gene in CE10 WT. Gene expression analyses were based off three biological repeats. * denotes $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$; **** $p \leq 0.0001$; ns (no significance) $p > 0.05$.

Results from a previous RNA-Seq experiment had revealed that in CFT073, genes involved in colibactin production were down-regulated upon exposure to D-serine (Connolly *et al.*, 2021). Colibactin is a bacterial genotoxin produced by some strains of *E. coli* and has been linked to CRC in humans. Interestingly, analysis of the RNA-Seq data presented in Chapter 3 for the pairwise comparison of CFT073 vs. Δ*dsdC* had revealed poor coverage for the colibactin biosynthesis region. As analysis of the ChIP-Seq data had revealed a peak at the intergenic

region between the colibactin master regulator, *clbR*, and a colibactin synthesis gene, *clbB*; RT-qPCR analysis was performed on *clbR* and another colibactin gene, *clbA*. RNA was extracted from CFT073 WT and Δ*dsdC* isogenic mutants. The samples were cultured in M9 minimal media for 3 hours and spiked with 1 mM D-serine and left to incubate for a further 2 hours. RT-qPCR analysis on three biological replicates revealed a 2.02-fold (*p*-value 0.0003) increase in gene expression of the colibactin regulator, *clbR*, in the Δ*dsdC* mutant background (Fig. 4-2A). Moreover, upon exposure to D-serine, *clbR* gene expression increased further to 4.58-fold (*p*-value 0.0003) in the Δ*dsdC* mutant background (Fig. 4-2A), indicating that in the presence of D-serine, DsdC represses the colibactin master activator. In the presence of D-serine, expression of *clbA* was also increased in the Δ*dsdC* mutant background (2.85-fold; *p*-value 0.0008) (Fig. 4-2B).

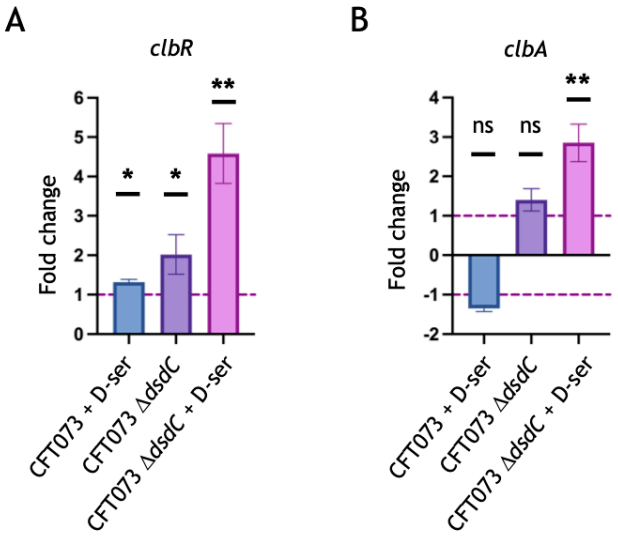


Figure 4-2 RT-qPCR analysis of colibactin genes. RT-qPCR analysis of two colibactin genes (A) *clbR* and (B) *clbA*, showing the levels of fold change in the CFT073 Δ*dsdC* mutant relative to the CFT073 WT. The purple dashed line indicates baseline expression of each gene in CFT073 WT. * denotes $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$; **** $p \leq 0.0001$; ns (no significance) $p > 0.05$.

4.3 Overexpression and purification of the LTTR, DsdC

4.3.1 Cloning and overexpression of DsdC with MBP and His tags

To purify DsdC, both an MBP tag and a 6xHis tag were used. Using the DNA sequence of *dsdC* (CFT073), *dsdC1* (CE10) and *dsdC2* (CE10), BamHI and HindIII restriction sites were amplified onto the 5' and 3' ends respectively. The product was then cloned on to the MBP in the pMAL-C5X vector, with the MBP tag at the N-terminal of *dsdC*. *dsdC1* (CE10) was also cloned onto a 6xHis tag on the pET-28a vector. The recombinant vectors were then cloned into the *E. coli* strain BL21 DE3 for overexpression.

Both MBP and 6xHis tag vectors rely on IPTG, a molecular mimic of allolactose, for activation of recombinant proteins. Small scale overexpression trials were performed with various concentrations of IPTG, time and temperature, to determine the best conditions for overexpression (Table 4-2). Western blot analysis and Coomassie staining revealed bands at ~78 kDa, which corresponded to the predicted molecular weight of DsdC (~36 kDa) and MBP tag (~42 kDa), and ~38 kDa for the His tagged DsdC1. It was established that 3 hours induction of 1 mM IPTG at 37°C was sufficient for DsdC overexpression (Fig. 4-3 A-D).

Temperature (°C)	Conc. of IPTG	Time after induction of IPTG (hours)
37	0.1mM	3
37	0.3mM	3
37	0.5mM	3
37	0.7mM	3
37	1.0mM	3
28	1.0mM	16
32	1.0mM	16

Table 4-2 Conditions used in overexpression trials of DsdC recombinant protein. To maximise overexpression of DsdC, samples were grown in LB to an OD_{600} of 0.6. The samples were then induced with several differing concentrations of IPTG, along with variations in temperature and exposure to IPTG.

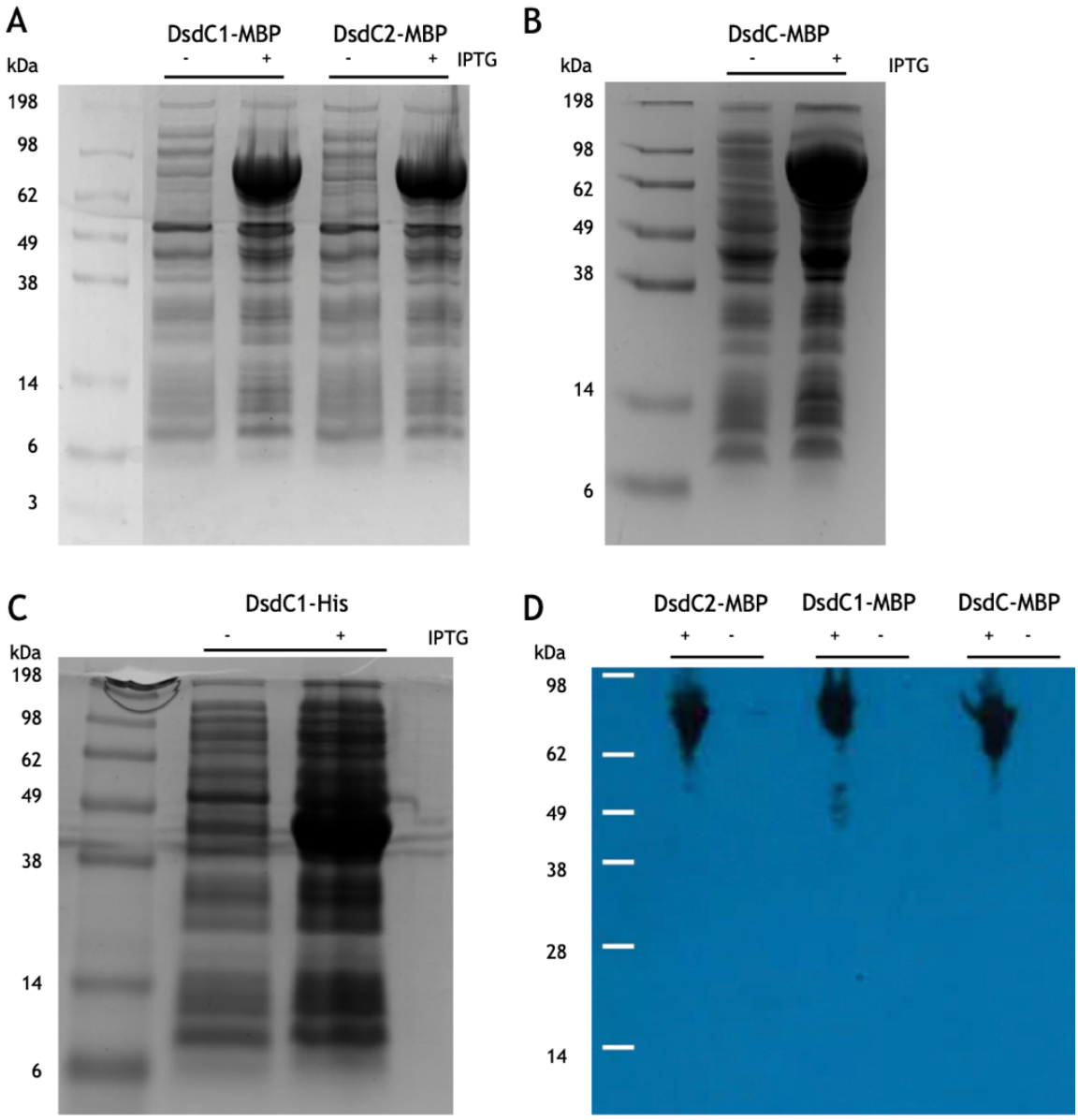


Figure 4-3 Coomassie Blue and Western blot analysis of the small-scale overexpression trials of DsdC. Overexpression trial of (A) pMAL-C5X_DsdC1 (CE10) and pMAL-C5X_DsdC2 (CE10) and (B) pMAL-C5X_DsdC (CFT073). "-" indicates no IPTG induction, "+" indicates induction of IPTG. Cell lysate indicated large bands between 62 and 98 kDa, with the MBP-DsdC fusion ~78 kDa. (C) Overexpression test of pET-28a-DsdC1 (CE10). "-" indicates no IPTG induction, "+" indicates induction of IPTG. Cell lysate indicated a band at ~38 kDa. (D) Western blot analysis of recombinant MBP-DsdC for CE10 DsdC2, CE10 DsdC1, and CFT073 DsdC, using anti-MBP antibody. "-" indicates no IPTG induction, "+" indicates induction of IPTG. Analysis indicated a band between 62 and 98 kDa, with the MBP-DsdC fusion ~78 kDa.

4.3.2 Purification of the recombinant DsdCs

Using the information obtained from the small-scale overexpression trials detailed above, full scale recombinant protein overexpression and purification of MBP-DsdC and His-DsdC was then performed. Cells were grown in 1 L of LB and induced with 1 mM IPTG. Cells were pelleted and resuspended in either MBP or His binding buffer and sonicated. The supernatant was then purified using affinity chromatography in either a MBP HP column or a HisTrap HP column. The elution fractions were then run on an SDS-PAGE gel to determine which elution fraction contained the purified recombinant protein (Fig. 4-4). Analysis of the gels revealed that DsdC purified in as a monomer when attached with an MBP-tag but potentially as a monomer and as a dimer when attached with a His-tag, as indicated by the red arrows. However, column purification is a crude method of protein purification and there were other bands present on all of the gels.

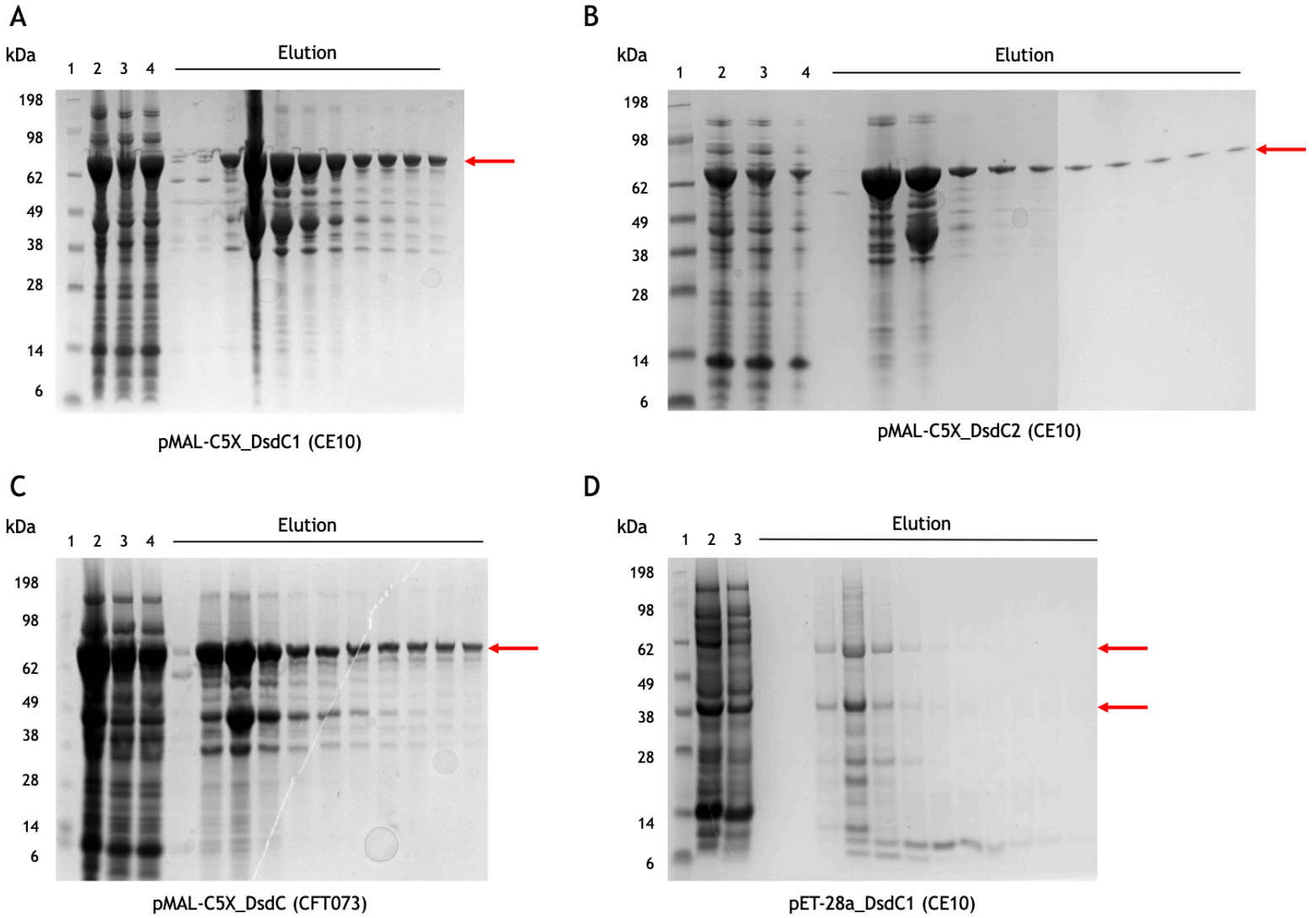


Figure 4-4 Purification of DsdC. SDS-PAGE analysis of purified DsdC, with the red arrows indicating the predicted size for DsdC-tagged

complexes. Lane 1, SeeBlue +2 ladder; lane 2, whole cell lysate; lanes 3-4, flow-through; lanes 5-15, elution fractions. The large band at ~78 kDa indicates (A) pMAL-C5X_DsdC1 (CE10) complex; (B) pMAL-C5X_DsdC2 (CE10) complex; (C) pMAL-C5X_DsdC (CFT073) complex. Purification of the (D) 6xHis-DsdC1 (pET28a_DsdC1) complex revealed two oligomeric states of DsdC1 at ~38 kDa and ~62 kDa.

4.3.3 Cleavage of MBP-DsdC tag

The MBP tag was used in purification to alleviate the potential solubility issue of DsdC. However, it is a large protein complex that could potentially interfere with DsdC's protein structure and function (Waugh, 2011). Therefore, Factor Xa was used for a trial cleavage experiment, performed on pMAL-C5X_DsdC1 to remove the MBP tag (Fig. 4-5). The trial cleavage experiment revealed that after 24 hours of digestion with Factor Xa, the recombinant protein was almost fully digested, leaving the singular MBP and DsdC proteins. However, upon cleavage with Factor Xa, the proteins precipitated out of solution, forming aggregates, suggesting that MBP was required for the solubility of DsdC.

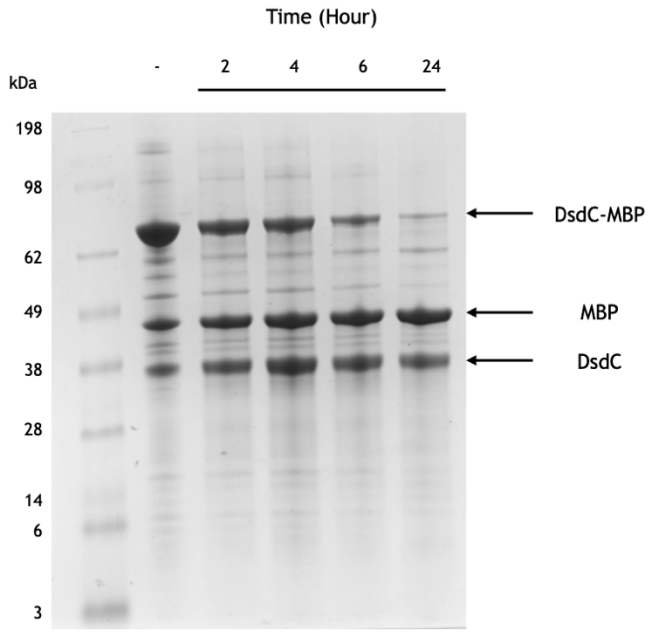


Figure 4-5 Cleavage of MBP tag from DsdC. SDS-PAGE showing the trial cleavage experiment of MBP from DsdC1 using Factor Xa over 24 hours. The black arrows indicate the DsdC-MBP complex and the singular MBP and DsdC proteins.

4.4 Validating DsdC binding through electrophoretic mobility shift assays (EMSA)

The ability of DsdC to bind to regions of the CFT073 and CE10 genome, as indicated by the ChIP-Seq data, was assessed using purified DsdC protein and the EMSA technique. EMSAs detect protein-DNA interactions *in vitro* and relies on the observation that protein-DNA complexes run slower in electrophoresis than free nucleic acid (Hellman & Fried, 2007). Regions of the CFT073 and CE10 genome that had been bound by DsdC in the ChIP-Seq experiment were amplified by PCR and DIG-labelled using the DIG gel shift kit (Roche). The DNA regions were then mixed with increasing concentrations of purified DsdC (CFT073), DsdC1 (CE10), or DsdC2 (CE10). EMSA analysis confirmed that DsdC bound to its own promoter region, and DsdC1 and DsdC2 also bound to the LPS glycosyltransferase gene, *waaV* (Fig. 4-6 and 4-7).

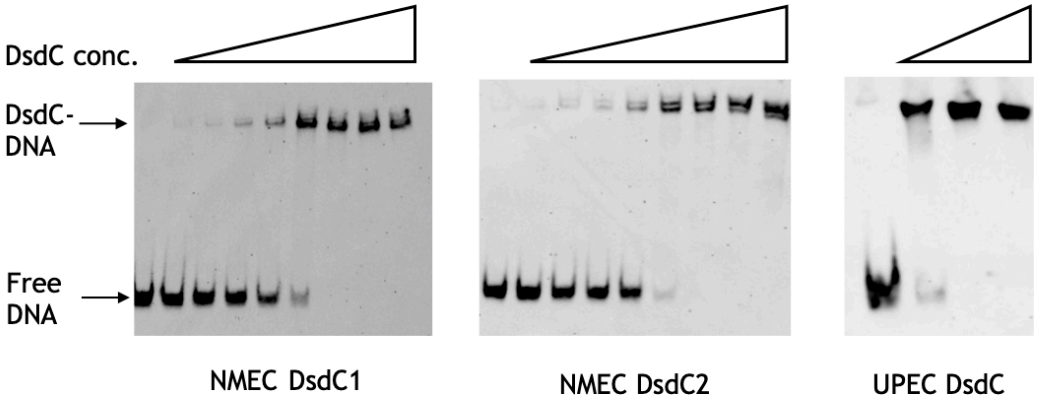


Figure 4-6 DsdC binds to the intergenic region upstream of the *dsdC* start codon. EMSA assay of DsdC1 (CE10), DsdC2 (CE10) and DsdC (CFT073) binding to the *dsdC-dsdX* intergenic region. Increasing concentrations of DsdC were used, as indicated by the triangle above the gel. DNA retardation was observed with the DsdC-DNA complexes remaining at the top of the gel whilst the free DNA ran further down the gel. EMSA experiments were performed in duplicate, with similar results obtained each time.

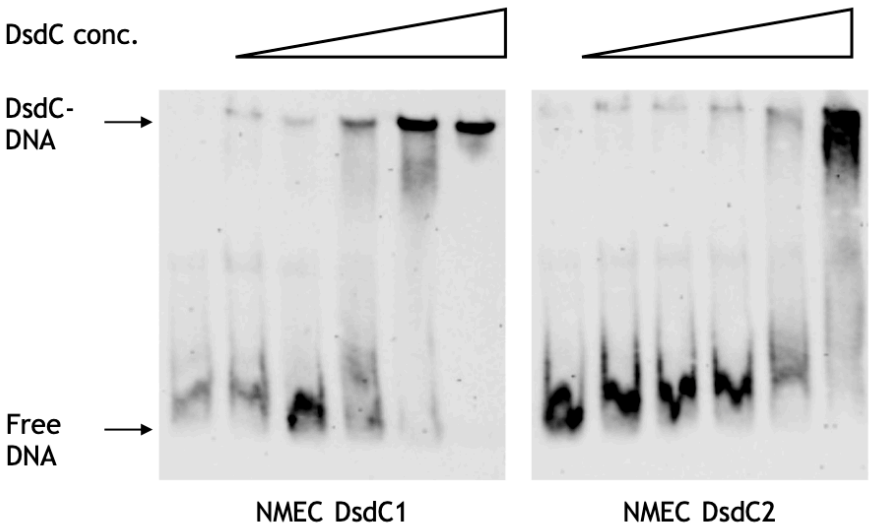


Figure 4-7 DsdC binds to the region upstream of the *waaV* start codon. EMSA assay of DsdC1 (CE10) and DsdC2 (CE10) binding to the *waaV-waaW* region. Increasing concentrations of DsdC were used, as indicated by the triangle above the gel. DNA retardation was observed with the DsdC-DNA complexes remaining at the top of the gel whilst the free DNA ran further down the gel. EMSA experiments were performed in duplicate, with similar results obtained each time.

These results correlated with previous findings that DsdC was able to bind to its own promoter, typical of a LTTR (Nørregaard-Madsen *et al.*, 1995). DsdC1 and DsdC2 were also shown to bind to the LPS glycosyltransferase gene, *waaV*. Together these data implied that DsdC was able to bind to regions of the genome outside of the *dsdC-dsdXA* intergenic region, validating the ChIP-Seq. Strikingly however, no other sites were able to be validated by EMSA. Indeed, multiple regions identified from the ChIP-Seq were tested including *espX4*, *kpsF, neuO, fimB, clbR, neuB, gadB,* and *envR,* however no shifts in DNA were observed in the presence of DsdC (Fig. 4-8). Both cleaved and intact MBP-DsdC recombinant proteins were tested, and 6xHis-DsdC, however no binding was observed.

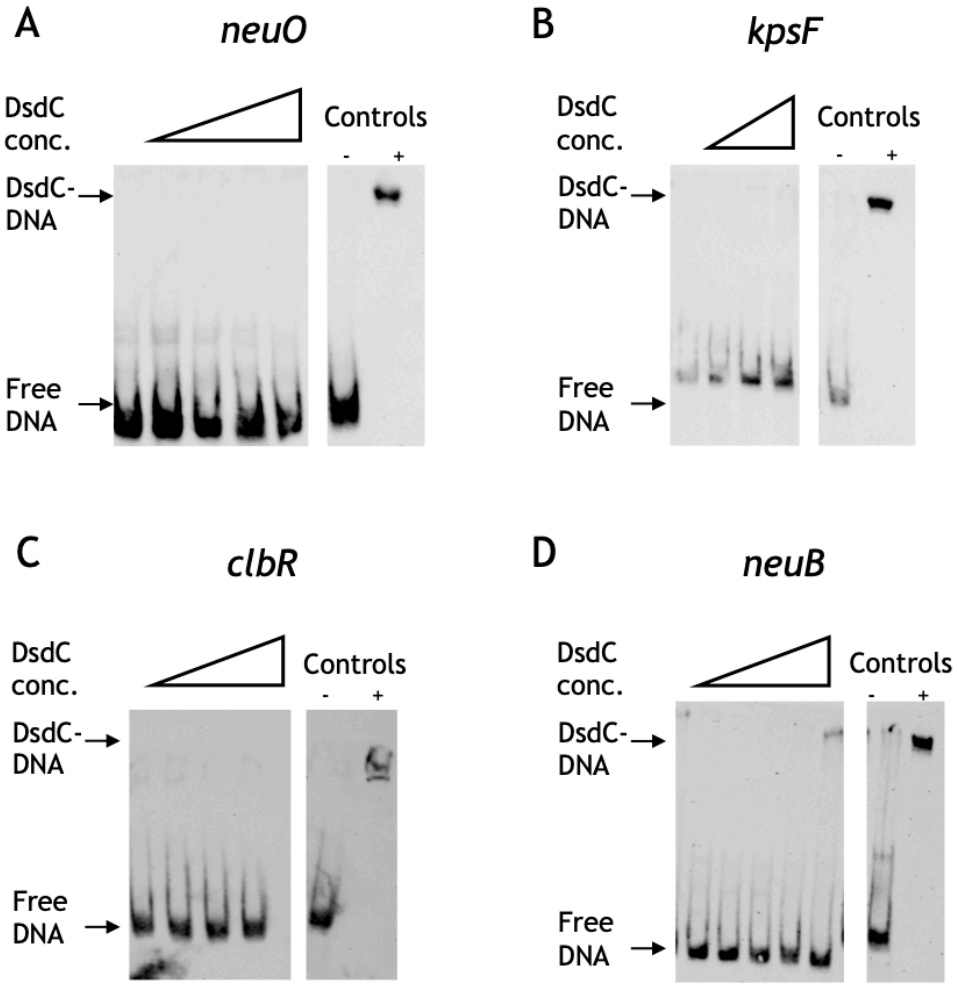


Figure 4-8 EMSAs did not show binding of DsdC to ChIP-Seq identified regions. (A) EMSA assay showing no DsdC binding to the predicted DsdC binding site for *neuO*. Increasing concentrations of DsdC were used, as indicated by the triangle above. (B) EMSA assay showing no DsdC binding to the predicted DsdC binding site for *kpsF*. Increasing concentrations of DsdC were used, as indicated by the triangle above. (C) EMSA assay showing no DsdC binding to the predicted DsdC binding site for *clbR*. Increasing concentrations of DsdC were used, as indicated by the triangle above. (D) EMSA assay showing no DsdC binding to the predicted DsdC binding site for *neuB*. Increasing concentrations of DsdC were used, as indicated by the triangle above. A positive DNA retardation control was used in the form of the *dsdC-dsdX* intergenic DNA and a negative control of using *ctx* DNA, a gene absent in *E. coli*. EMSA experiments were performed in duplicate, with similar results obtained each time.

4.5 Elucidating the binding motif of DsdC

Although the resolution of ChIP-Seq is high and gives vast amounts of information regarding potential binding sites, it does not give sufficient resolution to resolve the binding motif at the nucleotide level. DNase I footprinting is a technique that detects protein-DNA interaction by subjecting DNA to digestion by DNase I. DNA that is bound by the protein, will be protected from digestion, thereby revealing the specific binding motif of the protein (Galas & Schmitz, 1978). As discussed in Chapter 1, LTTRs typically bind as dimers in two distinct places in the promoter region; at a regulatory binding site (RBS) -35 to +20 bp and at an activation binding site (ABS) -40 to -20 bp (Fig. 4-9) (Goethals *et al.*, 1992; McFall *et al.*, 1998). They often bind to a palindromic DNA sequence, referred to as the LTTR box, consisting of a T-N_{11}-A sequence, which is present at the RBS site but not the ABS site. To elucidate the binding motif of DsdC, the *dsdC-dsdX* intergenic region was radiolabelled and using purified MBP-DsdC, a DNase I footprint was performed. This was performed at the University of Birmingham in conjunction with Dr Douglas Browning, who completed the experiment.

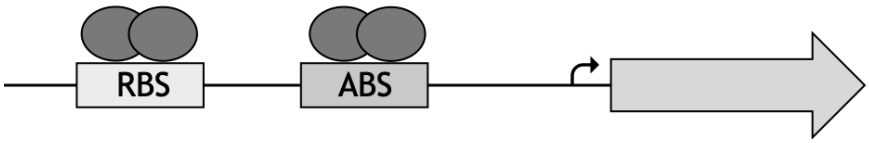


Figure 4-9 A graphical representation of LTTR binding. LTTRs typically bind as dimers in two distinct places within a promoter region, in the RBS (regulatory binding site) and the ABS (activation binding site), adapted from Maddocks & Oyston, 2008.

The DNase I footprint revealed two distinct binding sites of DsdC within the *dsdC-dsdX* intergenic region, as indicated in Figure 4-10. The first protected site was a 19 bp region that overlapped the predicted -10 site of *dsdC*, as predicted by RegulonDB. This site contained a putative T-N_{11}-A LTTR box. The second protected site was a 47 bp region that was closer to the start codon of *dsdC*. This site contained two putative LTTR boxes, a T-N_{24}-A LTTR box and a secondary T-N_{11}-A LTTR box further along the region. Strikingly, addition of D-serine to the footprint caused a reduction in binding to the second protected region. These results revealed the novel binding motif for DsdC and suggested that D-serine influences the strength of DsdC binding to target promoters.

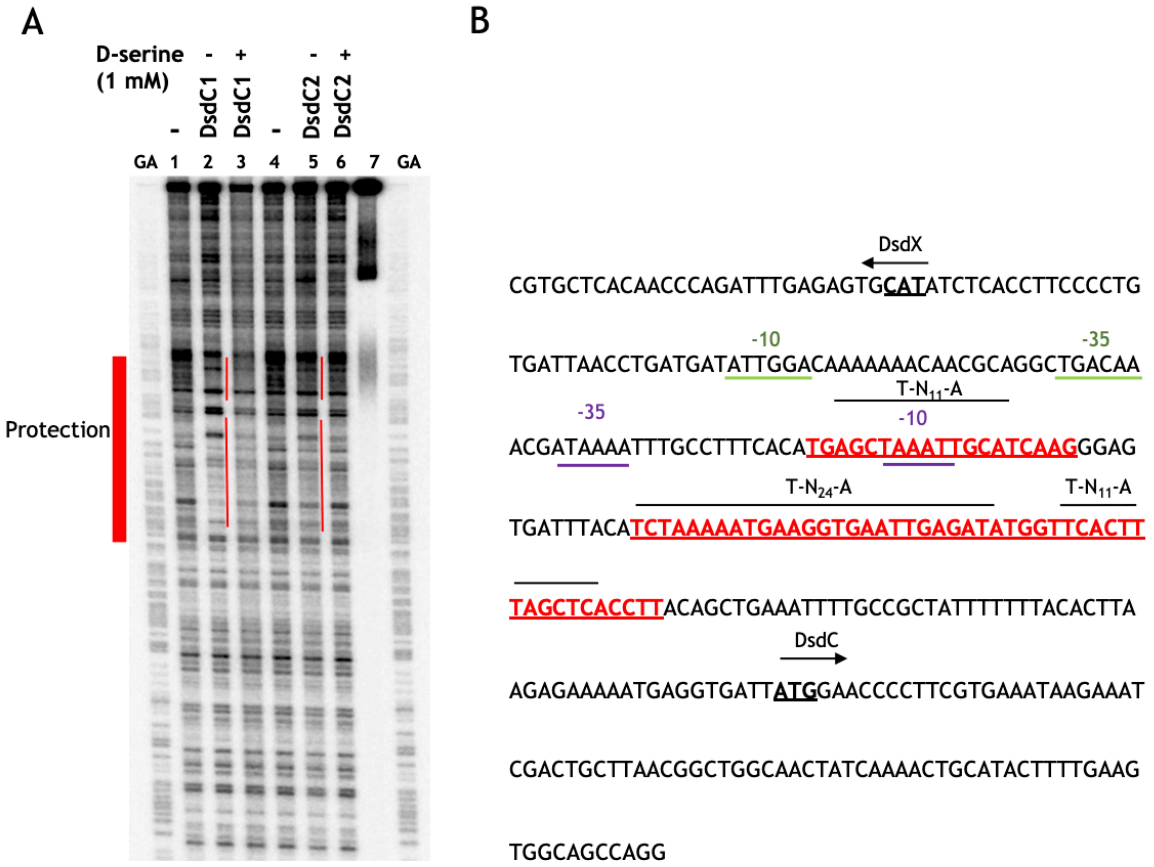


Figure 4-10 Elucidating the binding motif of DsdC. (A) DNase I footprinting analysis revealed the protected region (red) of the CE10 genome, where DsdC1 (CE10 derived) and DsdC2 (CE10 derived) were bound to the *dsdC1-dsdX1* region, in the presence and absence of D-serine (lanes 2-6). Lanes 1 and 4 indicate no protein controls and lane 7 indicates uncut fragment. (B) Organisation of the intergenic region between *dsdC* and *dsdX*. Nucleotides in bold signify the start codon of the gene above. Nucleotides underlined in green signify the -10 and -35 region of *dsdX* as reported by Nørregaard-Madsen, Mcfall and Valentin-Hansen, (1995). Nucleotides underlined in purple signify the -10 and -35 region of *dsdC* as predicted by RegulonDB. Nucleotides in bold and in red signify the protected region of the genome, revealing the binding motif of DsdC. Highlighted above the binding motifs are potential DsdC LTTR binding boxes of T-N_{24}-A and T-N_{11}-A.

4.6 Discussion

4.6.1 Differential gene expression of LPS and capsular biosynthesis genes validates the RNA-Seq data

The RNA-Seq data reported in Chapter 3 revealed there was differential expression in 30 genes in the pairwise comparison between CE10 WT and Δ*dsdC1/2*, upon exposure to D-serine. Strikingly, most of these genes were involved in capsular and LPS biosynthesis, two key virulence determinants of NMEC. RT-qPCR was used to measure the differences in transcript levels in four genes related to the capsule and LPS, and thus demonstrated the validity of the RNA-Seq data (Table 4-1). These results further indicated that DsdC plays a role in regulation of NMEC specific pathogenic traits, repressing expression of capsular and LPS biosynthesis genes, whilst activating expression of *neuO*, an *O*-acetyltransferase.

Furthermore, RT-qPCR analysis was carried out on two colibactin genes, *clbR* and *clbA*. Colibactin is a bacterial genotoxin that can be produced by UPEC strains and is synthesised by genes encoded on a 54-kb genomic island (Nougayrède *et al.*, 2006). *clbR* and *clbA* are transcribed divergently from the main operon, with *clbR* the master regulator of colibactin production (Wallenstein *et al.*, 2020). Examination of the *clb* region in the RNA-Seq data revealed poor coverage and so no analysis could be made in regard to differential expression of the colibactin synthesis genes between the CFT073 WT and Δ*dsdC*. RT-qPCR analysis of *clbR* revealed a 2-fold increase in gene expression between Δ*dsdC* mutant and WT, which further increased to 4.58-fold upon exposure of D-serine. This data implies that DsdC represses the colibactin synthesis activator in CFT073, suggesting a novel role for DsdC in regulation of a bacterial genotoxin.

The differential expression of the K1 capsular biosynthesis genes in CE10 and the colibactin synthesis genes in CFT073, in response to exposure to D-serine, further indicates that DsdC has been tailored for strain-specific lifestyles in ExPEC. Adapting to a specific niche within the host is a multifactorial process, often involving the alteration of carbon source intake and expression of virulence factors involved in host colonisation (reviewed in O'Boyle *et al.*, 2020).

Virulence factors are often encoded on genetic islands (Ho-Sui *et al.*, 2009), which is indeed the case for the K1 capsule and colibactin synthesis genes which reside on the *kps* and *pks* islands respectively (Nougayrède *et al.*, 2006; Silver & Vimr, 1990). As genetic islands are often acquired through HGT, they must be integrated into a cells already established regulatory network (Perez & Groisman, 2009b). Ancestral TFs are therefore often repurposed and tailored to accommodate the newly integrated genomic islands (Perez & Groisman, 2009a; Connolly *et al.*, 2019; reviewed in O'Boyle *et al.*, 2020). As metabolism and virulence gene expression are two closely linked processes within the bacterial cell (reviewed in Poncet *et al.*, 2009), hypothetically exposure to D-serine in the urinary tract and the brain, mediated through DsdC, may be used as signal to alter the regulation of colibactin and capsular biosynthesis, two key virulence factors of ExPEC. This is an interesting hypothesis which further chapters in this thesis will try and elucidate.

4.6.2 DsdC shows direct binding to *dsdCX* and to the LPS glycosyltransferase gene, *waaV*

DsdC is a LTTR, which in the past have generally been reported to be insoluble proteins (Ezezika *et al.*, 2007). A common method to overcome protein insolubility is by using a fusion protein strategy. This involves cloning a gene of interest onto a plasmid that will increase the solubility. These plasmids often contain the genes encoding glutathione S-transferase (GST), thioredoxin A or MBP (Kapust & Waugh, 1999; Young *et al.*, 2012). To purify DsdC, both an MBP and a 6xHis tag was used. Small-scale overexpression trials revealed that 1 mM IPTG was sufficient for overexpression of the recombinant protein, with both recombinant proteins appearing to be overexpressed at similar levels.

The ChIP-Seq data reported in Chapter 3 revealed that DsdC bound to multiple regions of the CE10 and CFT073 genome. Indeed, it was shown that in CE10 DsdC bound to 217 regions and in CFT073 DsdC bound to 129 regions (Table 3-3). To validate the ChIP-Seq data, purified DsdC was used in EMSAs to detect binding to predicted regions. DsdC1 and DsdC2 (CE10) were shown to bind to the region upstream of the *waaV* gene and to the intergenic region between *dsdCX*, but not to any of the other ChIP-Seq predicted sites. DsdC (CFT073) was only shown to bind to the intergenic region between *dsdCX*. Although the ChIP-Seq data had

indicated multiple regions of DsdC binding, there are several explanations for why these were not evident in the EMSAs. Indeed, EMSAs are *in vitro* techniques and therefore environmental components required for native binding may have been absent. Additionally, LTTRs often require a co-inducer in order to bind (reviewed in Maddocks & Oyston, 2008), and although D-serine was tested in the EMSAs and did not appear to elicit an effect, the concentration may have been insufficient for facilitating *in vitro* binding. Further as TFs sometimes work in tandem with other TFs (Martínez-Antonio & Collado-Vides, 2003), potentially *in vitro* binding was not evident because a secondary TF was required. Indeed, it has been reported that DsdC and catabolite repressor protein (CRP) can co-regulate the *dsdXA* locus, with CRP presence increasing expression of the genes 7-fold (Nørregaard-Madsen *et al.*, 1995). Furthermore, although crude preparations of protein are routinely used in EMSAs, the protein quality could have been affected by various enzymes still in the preparation, affecting the efficiency of the binding (Holden & Tacon, 2011).

Conversely, although many of the ChIP-Seq binding sites were unable to be validated *in vitro*, DsdC binding was still shown in the upstream region of the *waaV* gene. Further, the RNA-Seq and RT-qPCR data revealed differential gene expression of *waaV*, indicating that DsdC directly regulates an LPS biosynthesis gene, a key virulence mechanism in NMEC. Suggested future studies to validate the ChIP-Seq could include testing a different solubility tag for DsdC overexpression, perhaps GST. Moreover, further steps could be performed in the purification of DsdC, such as size exclusion chromatography and post tag cleavage. These future studies may then be able to validate more of the ChIP-Seq results presented here.

4.6.3 Elucidating the DNA binding motif of DsdC using DNase I footprinting

Although the ChIP-Seq analysis revealed a wealth of information regarding DsdC binding, it did not reveal the specific binding motif of DsdC. LTTRs often function as tetramers, binding as dimers at the RBS and ABS site within the intergenic region upstream of the promoter (McFall *et al.*, 1998; Schell, 1993). Typically, LTTRs have been found to protect large regions of DNA, often between 50-60 bp (Muraoka *et al.*, 2003). LTTRs often bind to a palindromic DNA

sequence, termed the LTTR box, consisting of the sequence T-N_{11}-A, although this can vary both in size and nucleotide composition (Goethals *et al.*, 1992). Conversely, it has recently been suggested that LTTRs rely heavily on DNA shape rather than on nucleotide sequence when recognising target DNA (Dorman *et al.*, 2020). Strikingly, the DsdC footprint revealed two distinct protection sites in the intergenic region between *dsdC-dsdX*. The first protected site was 19 bp in length and contained a putative A-N_{11}-T box. This site was closer to the start codon of *dsdX* and could indicate the potential binding region of DsdC when regulating *dsdXA*. The second protected region was 47 bp in length, indicative of a tetramer binding. This region contained two putative LTTR boxes, T-N_{24}-A and T-N_{11}-A. This site was closer to the start codon of the *dsdC* gene and could indicate the binding region of DsdC to its own promoter. Indeed, this was further indicated when D-serine was added to the reaction. LTTRs are known to negatively auto-regulate themselves, with transcription occurring when the protein has been dissociated from the promoter, often because it is interacting with the co-inducer (reviewed in Maddocks & Oyston, 2008). D-serine addition to the footprint caused a reduction in binding to the secondary protected region, suggesting that DsdC had been dissociated from the promoter region. Therefore, using DNase I footprinting, the DsdC binding motif has been elucidated. It is a 47 bp binding region, likely indicating that DsdC binds as a tetramer, in conjunction with other LTTR reports.

4.6.4 Conclusion

To survive in fluctuating and competitive environmental conditions, bacteria maintain gene expression under tight regulatory control. Such is the importance of gene regulation, *E. coli* has devoted 6% of its genome to encoding gene regulatory elements and encodes up to 300 putative TFs (Martínez-Antonio & Collado-Vides, 2003). As discussed in Chapter 3, ChIP-Seq and RNA-Seq are both powerful global tools that can be used in an unbiased manner to elucidate a TFs regulon. These tools were used to elucidate the regulon of DsdC in CE10 and CFT073, revealing that DsdC bound in a global manner to multiple regions of the genome, and upon exposure to D-serine, affected the expression of 30 genes in CE10. In this chapter, using RT-qPCR, the RNA-Seq data was validated, demonstrating differential expression of LPS and capsular biosynthesis genes in the CE10 $\Delta\Delta$*dsdC1/2* background. Furthermore, differential gene expression of

the colibactin synthesis genes in the CFT073 Δ*dsdC* background was revealed, indicating a potential role of DsdC in colibactin synthesis. Using EMSAs, two of the ChIP-Seq binding peaks were validated and it was demonstrated that DsdC directly binds to an LPS glycosyltransferase gene, a key virulence factor in NMEC. Moreover, using DNase I footprinting, the binding motif of DsdC was elucidated, revealing a 47 bp region close to the *dsdC* start codon, indicating that DsdC binds as a tetramer. These results combined further highlight that DsdC plays a larger role in gene regulation beyond D-serine metabolism and further suggests that DsdC has been tailored in ExPEC strains for strain-specific lifestyles.

Chapter 5 DsdC influences susceptibility of neonatal meningitis-associated *Escherichia coli* strain CE10 from K1 bacteriophages, through the regulation of the capsular *O*-acetyltransferase gene, *neuO*

5.1 Introduction

Gene expression is a tightly regulated process in bacteria, often governed by environmental signals and cues. In Chapters 3 and 4, the novel roles of DsdC in CFT073 and CE10 gene regulation were discussed. Intriguingly, it was discovered that DsdC bound and altered gene expression of capsular and LPS biosynthesis genes in CE10 in the presence of the host metabolite, D-serine. The LPS and capsule are a bacterium's first line of defence against the host immune system, and thus are deemed two key virulence factors in a bacterium's arsenal. There are over 70 distinct capsular (K) antigens found in *E. coli* (Jann & Jann, 1992), however, the K1 antigen is often over-represented in NMEC infection (Robbins *et al.*, 1974). The K1 capsular polysaccharide is comprised of Neu5Ac polysialic acid, and has been reported to play a role in NMEC's ability to evade phagocytosis and enhance serum resistance (reviewed in Silver & Vimr, 1990; Croxen & Finlay, 2010). Furthermore, molecular mimicry of the host polysialic acid of N-CAM, results in the poor immunogenicity and enhanced survival of the K1 capsule strains within the host (reviewed in Silver & Vimr, 1990). Moreover, the K1 capsule is also phase variable, with the polysialic acids that comprise the capsule able to *O*-acetylate (Orskov *et al.*, 1979).

Strikingly, irrespective of D-serine presence, *neuO* was one of the most downregulated genes in the CE10 Δ*dsdC1/2* transcriptomics in relation to the CE10 WT. NeuO is a capsular *O*-acetyltransferase that was first described in *E. coli* in 2005 by Deszo *et al.*, (2005). *O*-acetylation of capsular polysaccharide is a mechanism used by many bacteria and fungi including *Streptococcus pneumoniae*, *Neisseria meningitidis* and *Cryptococcus neoformans* (reviewed in O'Meara & Alspaugh, 2012; Geno *et al.*, 2015; Tzeng, Thomas & Stephens, 2016). Considerable research has been performed in *N. meningitidis* where *O*-acetylation has been shown to be involved in protection from serum bactericidal activities, immunogenicity and physiology (Fusco *et al.*, 2007), indicating *O*-acetylation plays a role in virulence in *N. meningitidis*. Conversely, in *E. coli* the role of capsule *O*-acetylation has been less clear, with one study reporting *O*-acetylated strains had an increased virulence in patients, whereas another study revealed no difference in virulence between *O*-acetylated and non-*O*-acetylated strains in a neonatal mouse model of infection (Colino & Outschoorn, 1999; Frasa *et al.*, 1993). Further findings have suggested that *O*-acetylation may play more

of an environmental role, with strains that were *O*-acetylated enhancing desiccation resistance (Mordhorst *et al.*, 2009).

neuO, the *O*-acetyltransferase in *E. coli*, was originally described as being encoded on the prophage CUS-3, which is integrated in the NMEC genome in the hypervariable region between *dsdC-argW* (Deszo *et al.*, 2005). NeuO-mediated *O*-acetylation is phase variable and is modulated by slipped strand mispairing (SSM) of 7 highly repeated nucleotides (5'-AAGACTC-3'). This contingency locus occurs at the start of the ORF of *neuO*, in a region termed the polyΨ region (Fig. 5-1). Multiples of 3 heptanucleotides will result in an active protein; whereas loss or gain of anything other than 3 heptanucleotides will result in a translational frame shift, and in an inactive truncated protein. This chapter will discuss the role of DsdC in the regulation of the capsular *O*-acetyltransferase, *neuO*, and any phenotypes observed from this regulation using a K1 bacteriophage killing assay.

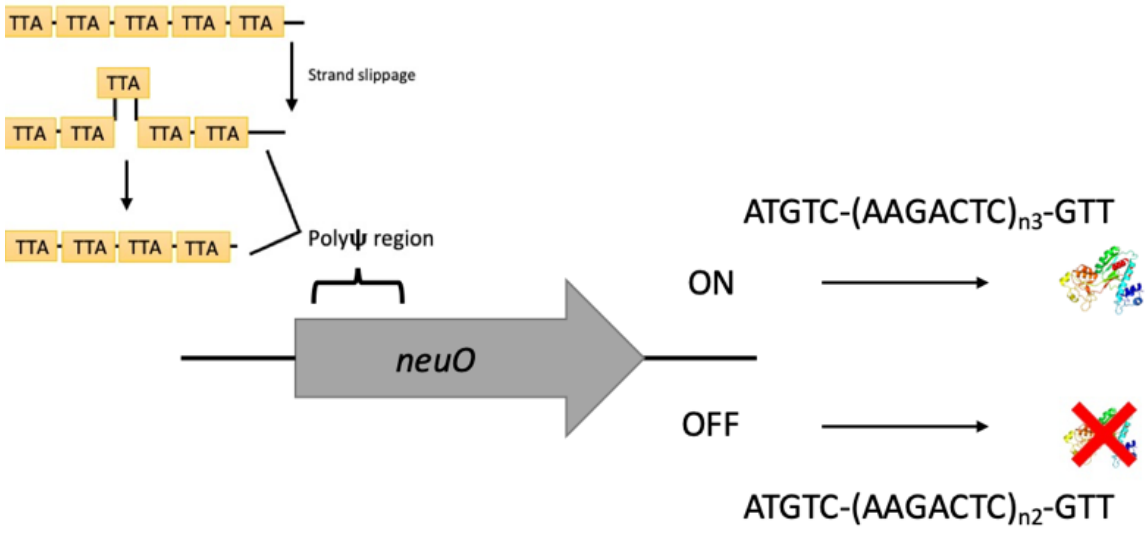


Figure 5-1 Phase variability of *neuO*. The phase variability of *neuO* is mediated by slipped strand mispairing of DNA during DNA replication. This often occurs in bacteria within regions of highly repeated nucleotides, with an example shown as the inset. This contingency locus occurs in the polyΨ region of *neuO*, at the start of the ORF. Multiples of 3 heptanucleotides (5'-AAGACTC-3') result in an active protein. Loss or gain of anything over than 3 heptanucleotides results in an inactive, truncated protein.

5.2 The binding dynamics of DsdC on the *neuO* gene

Comparative transcriptomics had revealed *neuO*, an *O*-acetyltransferase, was one of the most down-regulated differentially expressed genes in the CE10 data sets, irrespective of exposure to D-serine (Fig 3-23). Analysis of the CE10^{DsdC} ChIP-Seq binding profiles revealed that DsdC bound in the ORF of *neuO* (Fig. 5-2A). There also appeared to be a secondary peak, indicated by the gold dashed line on Figure 5-2(A), at the end of the gene. However, upon further analysis, the origin of the reads were unable to be confirmed as specific to that region.

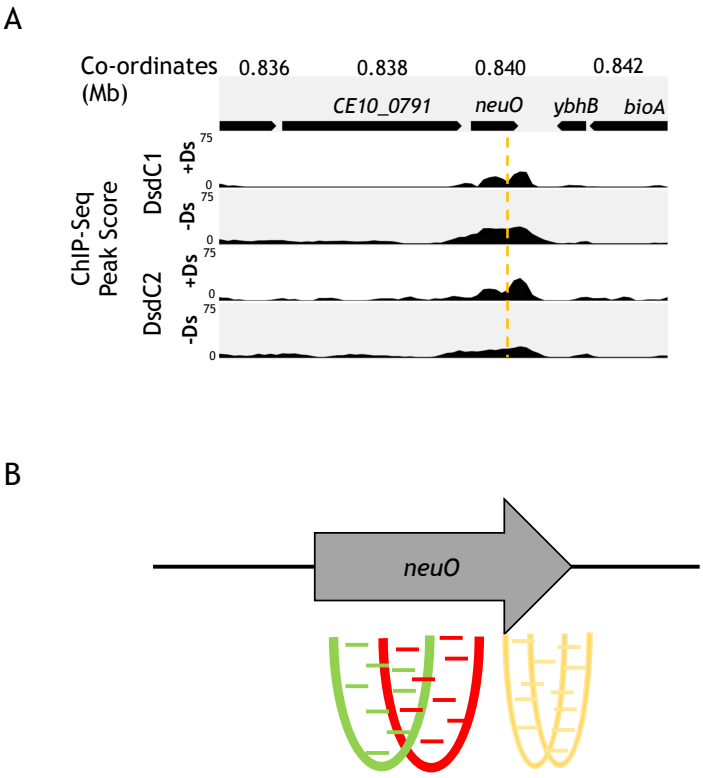


Figure 5-2 *neuO* binding dynamics. (A) DsdC1 and DsdC2 ChIP binding peaks for *neuO*, in the presence and absence of D-serine, are shown to be intragenic. The gold dashed line indicates the two binding peaks of DsdC for *neuO*. (B) A graphical representation of the two binding peaks of *neuO*. The peak to the left of the gold dashed line showed single reads mapping in both the forward (green) and reverse (red) direction. The peak to the right of the gold dashed line showed non-specific read mapping (yellow), indicating it matched to somewhere else on the genome and the software was unable to clarify which was the true peak.

5.3 NeuO is not carried on the mobile genetic element, CUS-3, in NMEC CE10

Original reports stated that *neuO* was encoded on the prophage CUS-3, and was integrated within the RS218 genome, a prototypical NMEC K1 strain, in the hypervariable region between *dsdC-argW* (Deszo *et al.*, 2005). Intriguingly, analysis of the CE10 genome revealed that *neuO* was encoded 2 mb upstream of *dsdC* (Fig. 3-14). Nucleotide BLAST analysis of the 40 kb CUS-3 genetic island from the RS218 genome against the CE10 genome revealed 38 out of the 59 genes that encode CUS-3 had no similarity to any region in the CE10 genome.

Using PHASTER, a database that identifies prophage sequences within the genome (Arndt *et al.*, 2016), 15 prophage regions were identified in the CE10 genome. This included a 16.2 kb region immediately upstream of *neuO*, between 824561 and 840854 bp. This region was shown to carry an incomplete prophage of 15 genes. The incomplete prophage found had homology to the bacteriophage HK620. Using EasyFig (Sullivan *et al.*, 2011), a linear BLAST comparison was made between CUS-3 in RS218 and the incomplete prophage in CE10 (Fig. 5-3). This analysis revealed homology at the start of the CUS-3 prophage and the incomplete prophage between *neuO* and the endosialidase gene. Further homology was revealed at the end of the CUS-3 prophage and the incomplete prophage, between several hypothetical proteins. However, there was no similarity between 37 genes in the central region of CUS-3. This suggests that if CUS-3 originally carried *neuO* into CE10, the phage has since been excised, although *neuO* has been retained. These results therefore indicate that *neuO* in CE10 is no longer linked to the prophage CUS-3.

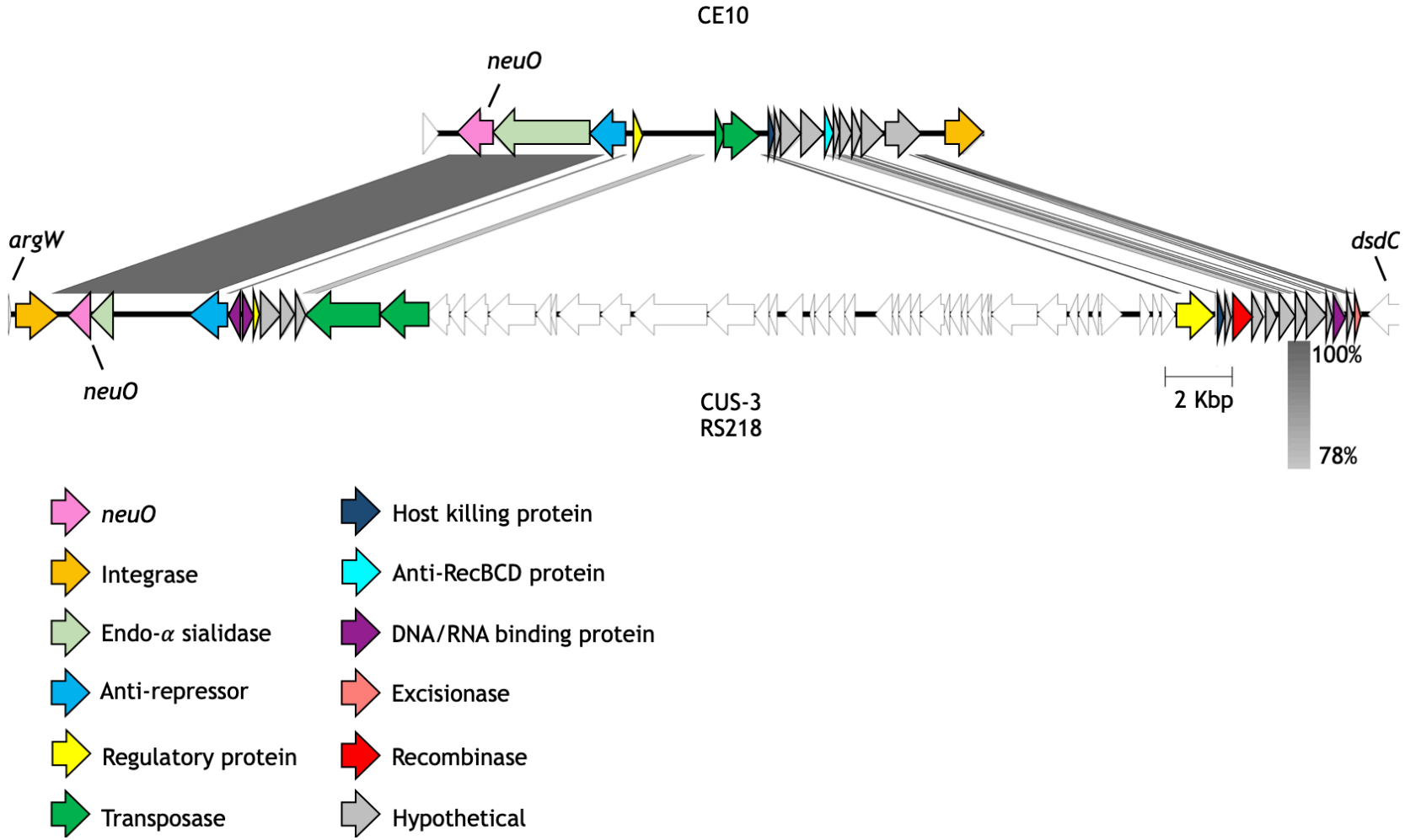


Figure 5-3 *neuO* is not encoded on CUS-3 in the CE10 genome. Linear BLAST analysis using EasyFig revealed little to no homology over the central region of CUS-3 to the incomplete prophage of CE10, indicating CUS-3 has been lost, or was never present. Shaded blocks between sequences represent homologous regions, shaded according to BLASTn. Coloured arrows indicate GO functionality.

5.4 Δ*neuO* does not affect growth of CE10

In order to establish if any future phenotypes found in the Δ*dsdC* mutants were mediated via NeuO, a Δ*neuO* mutant was generated in CE10 using the Lambda Red mutagenesis system, and provided by Dr James Connolly. To ensure there were no growth defects with the Δ*neuO* mutant that could contribute to any phenotypes observed, a growth curve was performed in M9 minimal media, in the presence and absence of D-serine, and qualitatively assessed (Fig. 5-4). The Δ*neuO* mutation appeared to have no effect on growth in M9 minimal media in comparison to the WT. *neuO* was also cloned into pACYC184 and transformed into Δ*neuO*, with the growth also comparable to the WT.

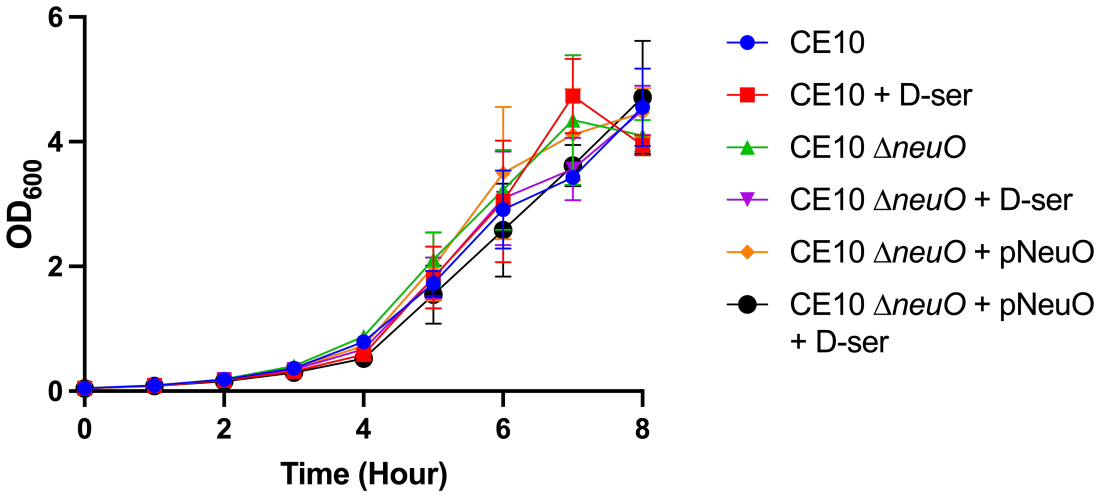


Figure 5-4 Growth profiles of WT and *ΔneuO* mutant in M9 minimal media, in the presence and absence of D-serine. Growth curves of CE10 and isogenic mutants in M9 minimal media in the presence and absence of 1mM D-serine, measured as OD_{600} over time. Growth curves shown represent mean of triplicate experiments with error bars indicating SEM.

5.5 NeuO transcription is regulated via DsdC and mediates protection against K1 bacteriophages

The K1 capsule is a known barrier to many bacteriophages, including the lytic T7 bacteriophage (Scholl *et al.*, 2005). K1-specific phages, however, often carry endosialidases which have the ability to degrade the α2,8-linked polysialic acid capsule (Stummeyer *et al.*, 2006). As the ChIP and RNA-Seq data had indicated

that DsdC was regulating *neuO*, it was hypothesised that this may affect the capsule and therefore the ability of the K1 bacteriophage to bind and lyse the cell. As previously described in Goh *et al.*, 2017, a K1-specific bacteriophage was purchased from SSI Diagnostica, and the ability of the bacteriophage to lyse the bacterial cells was measured.

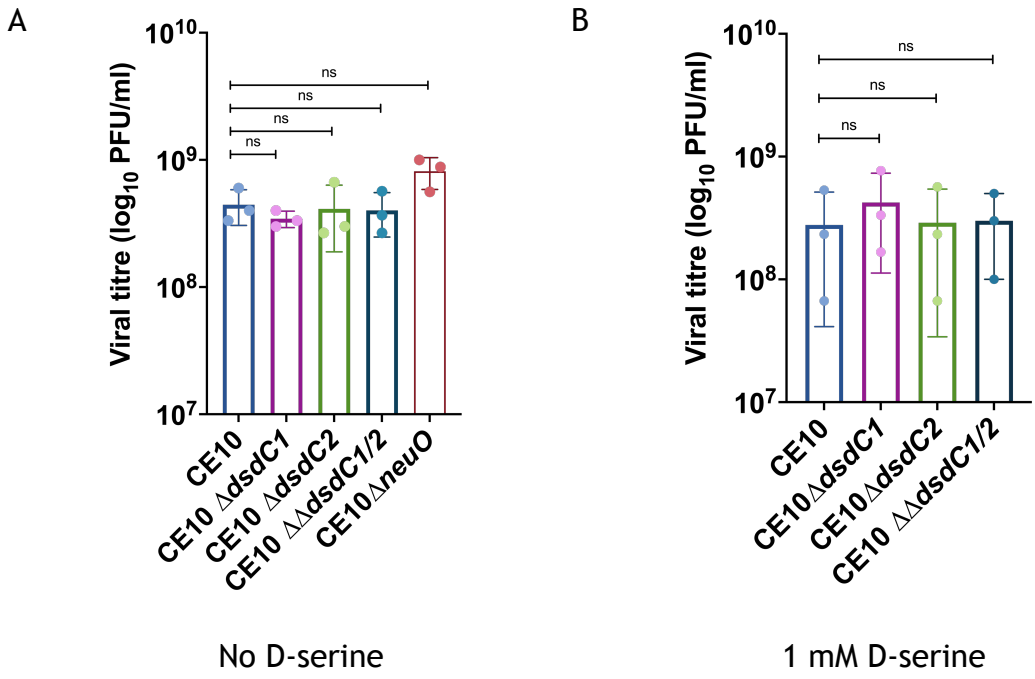


Figure 5-5 Quantification of viral titre in CE10 WT and isogenic mutants. The plaque forming units (PFU)/ml of the CE10 WT and isogenic mutants in (A) the absence of D-serine and (B) the presence of D-serine. The plaque assays were performed in triplicate with each replicate portrayed as a circle in the graph. Statistical non-paired Student's *t*-tests were performed on GraphPad Prism 8, comparing isogenic mutants against the CE10 WT. ns denotes a $p>0.05$.

Quantifying the viral titre showed the lytic effect the K1 bacteriophage was having upon CE10 and the Δ*dsdC* and Δ*neuO* isogenic mutants (Fig. 5-5). It was apparent from the PFU/ml, regardless of exposure to D-serine, there was no difference in the lytic activity of the K1 bacteriophage. It appears therefore that the K1 bacteriophage was lysing all of the bacteria comparably. However, although there was no difference observed in the phage titre results there was a difference in the size and morphology of plaques formed. Figure 5-6 shows the plaques formed on the CE10 WT and isogenic mutants and quantitative analysis of plaque sizes.

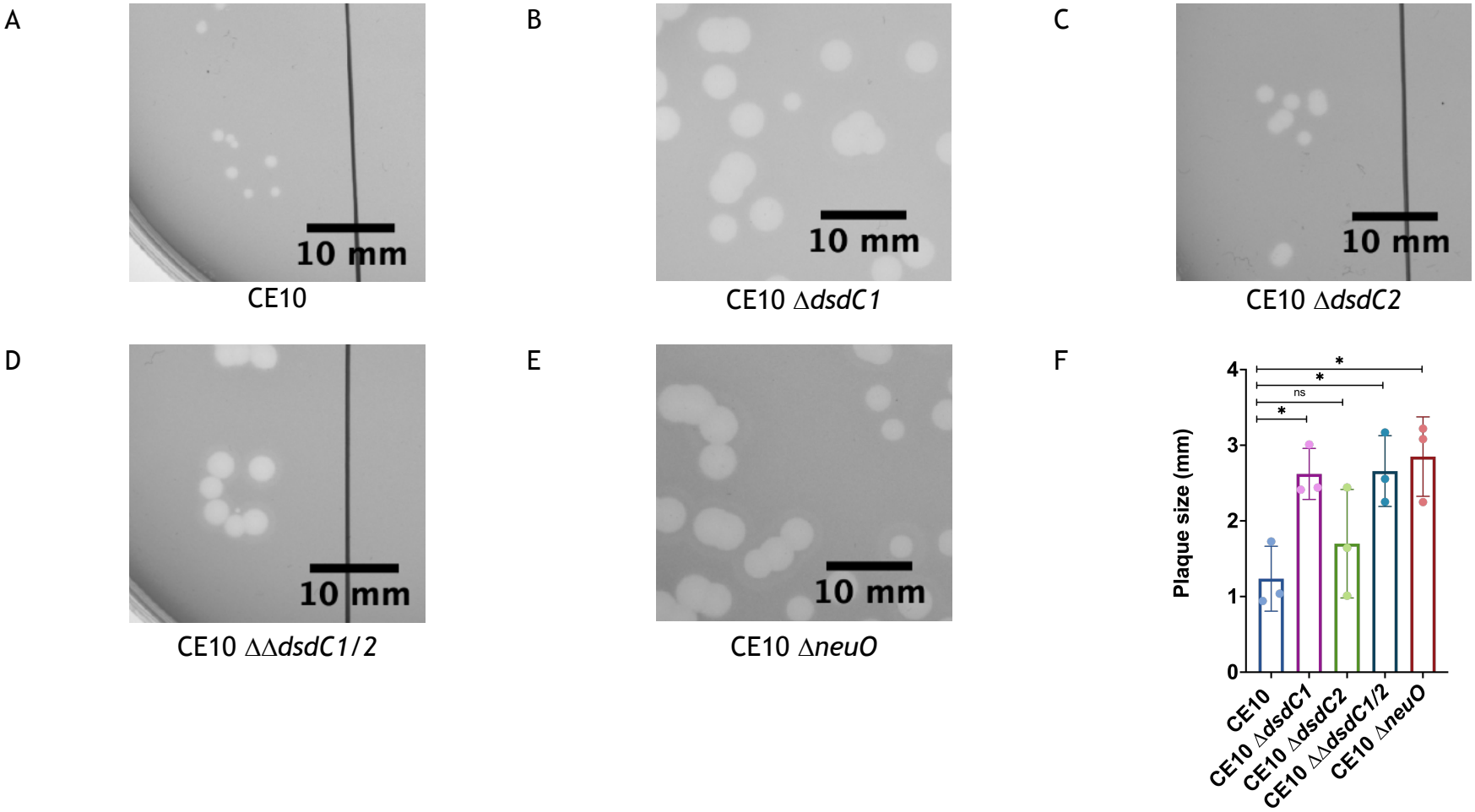


Figure 5-6 Δ*dsdC1* and Δ*neuO* present a large plaque phenotype. Images of phage plaques taken after infection on bacterial lawns of (A) CE10, (B) Δ*dsdC1*, (C) Δ*dsdC2*, (D) ΔΔ*dsdC1/2*, (E) Δ*neuO*. These images are representative of 3 independent experiments. Scale bars were generated on Fiji ImageJ software. (F) Quantitative analysis of plaque sizes measured in mm.

Statistical non-paired Student's *t*-tests were performed on GraphPad Prism 8, comparing isogenic mutants against the CE10 WT. * denotes a *p*-≤0.05; ns>0.05.

Significant differences in mean plaque size were observed comparing WT (1.15 mm) with Δ*dsdC1* (3.6 mm), ΔΔ*dsdC1/2* (3.65 mm) and Δ*neuO* (3.8 mm), but not Δ*dsdC2* (1.6 mm). This indicated that NeuO plays a role in protection from K1-bacteriophage induced lysis and that regulation of bacteriophage protection is likely DsdC1-specific. Further, there were also differences in the plaque morphology between samples. All of the plaques presented as clear, which is common for lytic phages, however some of the plaques formed on Δ*dsdC1*, ΔΔ*dsdC1/2* and Δ*neuO* mutants contained haloes (Fig. 5-7). These larger plaques with haloes surrounding them were not present on either the CE10 WT or Δ*dsdC2* plaques over 3 biological replicates. These results indicated a difference in the plaque phenotype presented between the WT and Δ*dsdC2* in comparison to Δ*dsdC1* and Δ*neuO*.

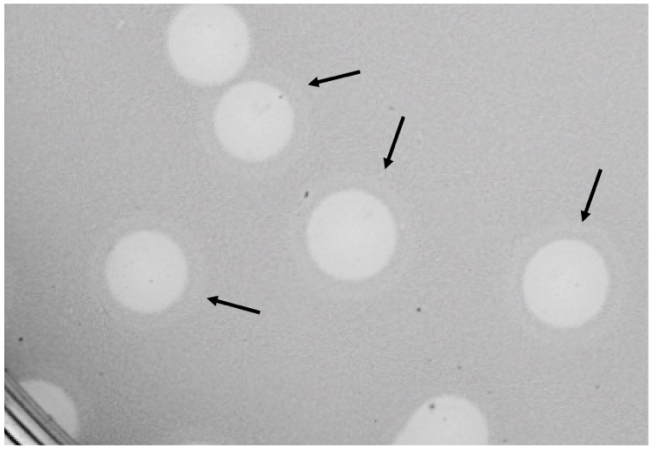


Figure 5-7 Representation of the haloes surrounding the plaques formed on Δ*neuO*. An image of haloes surrounding the plaques that were representative of plaques formed on Δ*dsdC1*, ΔΔ*dsdC1/2* and Δ*neuO* mutants. The black arrow signifies the halo.

The observation of a difference in phage plaque morphology suggested a difference in the K1 bacteriophage lytic lifestyle between WT and mutants (Abedon, 2021). A K1 bacteriophage killing assay, based on a method previously described in Goh *et al.*, 2017, was therefore performed to assess the rate of lysis

in CE10 and isogenic mutants. Briefly, the bacteria were grown on a microtitre plate for 3 hours in M9 minimal media, ensuring they were in logarithmic phase, as can be seen in Figure 5-8. 1 μl of K1 phage suspension was added to the samples and OD_{600} was measured every 30 minutes for 3 hours.

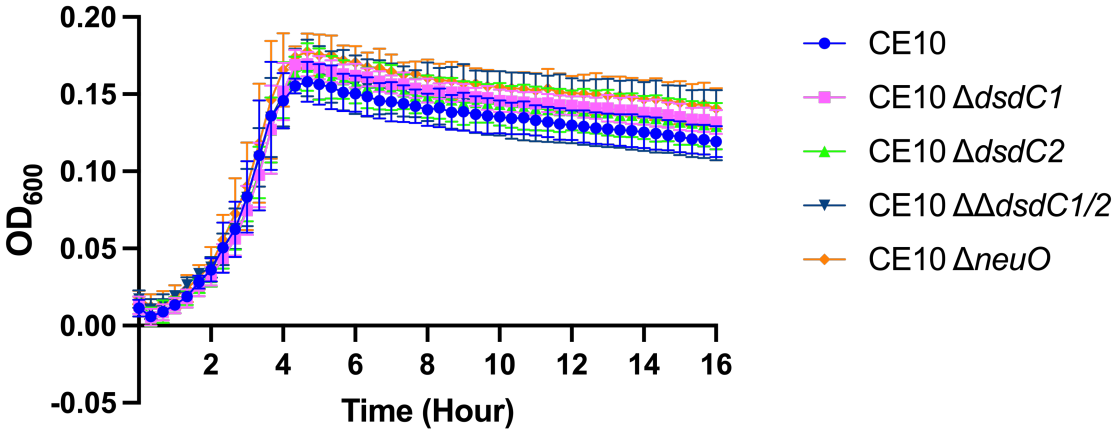


Figure 5-8 Growth profiles of WT and mutants in M9 minimal media, on a FLUOstar plate reader. Growth curves of CE10 and mutants in M9 minimal media, measured as OD_{600} over time on a FLUOstar Optima plate reader. Growth curves shown represent mean of triplicate experiments with error bars indicating SEM.

Figure 5-8 shows the growth profiles of CE10 and mutants on the FLUOstar Optima plate reader, measured over 16 hours. Qualitative assessment of Figure 5-8 revealed that the WT and mutants grew comparably, and therefore any effect observed in the K1 phage killing assay would be likely due to lysis by the bacteriophage. As lysis can start occurring quickly after addition of bacteriophage (Wang *et al.*, 1996), the OD_{600} was measured every 30 minutes for 3 hours (Fig. 5-9). Statistical non-paired Student's *t*-tests were performed at hour 4.5 (1.5 hours post infection (PI) of bacteriophage) and hour 6 (3 hours PI of bacteriophage).

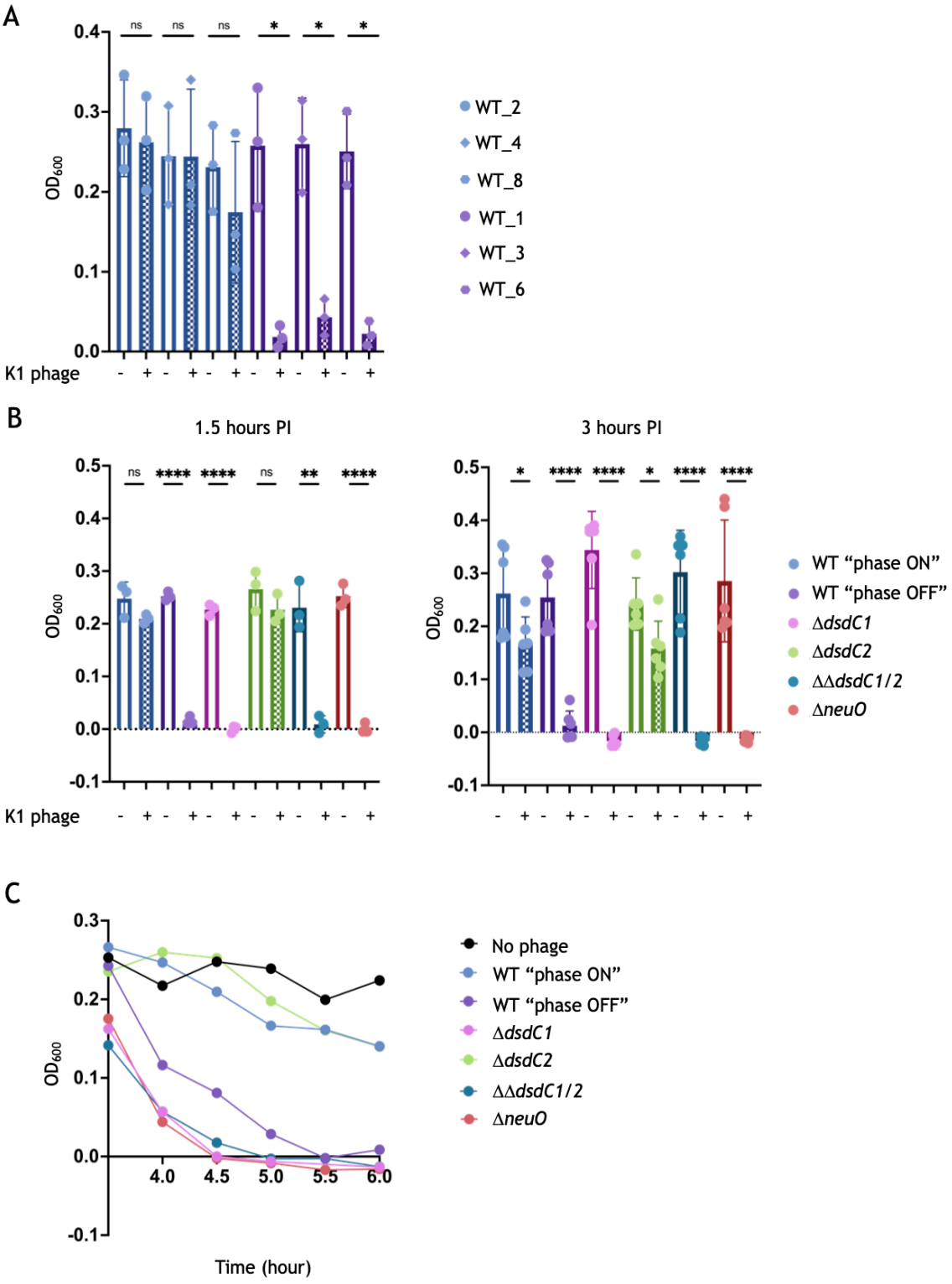


Figure 5-9 NeuO protects against K1 bacteriophage lysis, mediated via DsdC1. A K1 bacteriophage killing assay was performed on CE10 and isogenic Δ*dsdC* and Δ*neuO* mutants. (A) Two distinct WT phenotypes emerged in the K1 bacteriophage killing assay; WT colonies that could survive K1 bacteriophage-mediated lysis (blue) and WT colonies that couldn't (purple). OD_{600} of the samples measured 1.5-hours PI of K1 bacteriophage. (B) OD_{600} of the samples

measured 1.5- and 3-hours PI of K1 bacteriophage. (C) K1 bacteriophage killing assay measured as OD_{600} over time. The samples were measured every 30 minutes PI of K1 bacteriophage for 3 hours. Statistical non-paired Student's *t*-tests were performed using GraphPad Prism on (A) and (B), comparing infected samples to uninfected samples. * denotes $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$; **** $p \leq 0.0001$; ns $p > 0.05$.

Strikingly, the K1 bacteriophage killing assay revealed two distinct phenotypes from the clonal CE10 WT population (Fig. 5-9(A)); one set that was almost completely lysed, and one set that had no significant difference in killing in comparison to the non-infected sample. Through SSM of the polyΨ region, NeuO is phase variable. Therefore, NeuO can either be in a "phase ON" state and *O*-acetylating the capsule, or in a "phase OFF" state and no *O*-acetylation of the capsule occurs. Sequencing of the WT colonies revealed those that were associated with higher levels of protection from the K1 bacteriophage were NeuO "phase ON" and those that were almost entirely lysed were NeuO "phase OFF" (Table 5-1). These results indicated that the phase variation of NeuO mediates protection against K1 bacteriophage-mediated lysis.

Figure 5-9(B) shows the K1 bacteriophage killing results 1.5- and 3-hours PI. At 1.5-hours, significant differences were observed when comparing non-infected to infected samples in WT "phase OFF", Δ*dsdC1*, ΔΔ*dsdC1/2* and Δ*neuO* colonies. Conversely, Δ*dsdC2* infected with K1 bacteriophage showed no significant killing in comparison to the non-infected sample, nor did the WT "phase ON" colonies. At 3-hours PI, WT "phase OFF" colonies were almost entirely lysed and the Δ*dsdC1*, ΔΔ*dsdC1/2* and Δ*neuO* mutants were fully killed. Significant differences were also observed between non-infected and infected WT "phase ON" and Δ*dsdC2* colonies at this time point. Figure 5-9(C) shows the growth dynamics of CE10 and isogenic mutants in the K1 bacteriophage killing assay, measured every 30 minutes PI. Killing of the Δ*dsdC1*, ΔΔ*dsdC1/2* and Δ*neuO* mutants occurs within 30 minutes PI. By 4.5 hours they are almost completely lysed, with complete lysis occurring at hour 5. WT "phase OFF" colonies behave similarly to Δ*dsdC1* and are lysed by hour 5.5. WT "phase ON" colonies and the Δ*dsdC2* mutant were not killed as quickly and there was still survival at hour 6 (3 hours PI).

These data corresponded with the plaque size assays, which showed that Δ*dsdC1*, ΔΔ*dsdC1/2* and Δ*neuO* mutants all presented significantly larger plaques in comparison to the WT and Δ*dsdC2* mutant. These data combined potentially implies that *O*-acetylation of the capsule, mediated by NeuO, is used by the cell to stop K1 bacteriophage-mediated lysis. This protection is transient however, with the NeuO "phase ON" colonies starting to be lysed at 3-hours PI. These data also implies that *neuO* regulation is DsdC1-specific, with DsdC2 not eliciting the same "hyper-susceptibility" to K1 bacteriophage. To determine if it was a DsdC1 and NeuO mediated phenotype, pDsdC1 and pNeuO were transformed into Δ*dsdC1* and Δ*neuO*. Furthermore, as pDsdC1 and pNeuO contained their respective native promoter sequences, the GapA promoter region was also assembled onto the 5' end of *neuO* and cloned into pACYC184 for the K1 bacteriophage killing assay. This was performed in order to elicit a greater transcriptional response of *neuO*.

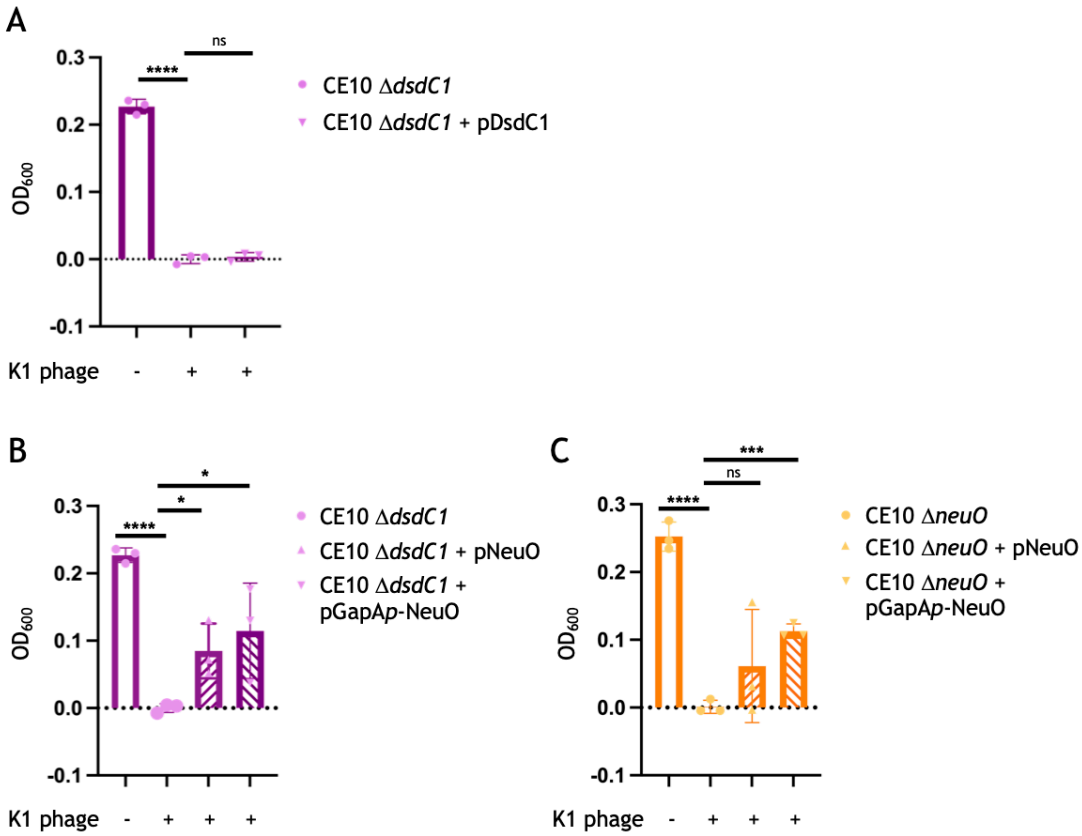


Figure 5-10 Complementation of NeuO restores protection against K1 bacteriophage-mediated lysis. A K1 bacteriophage killing assay measured as OD_{600}. (A) pDsdC1 transformed into Δ*dsdC1*, measured 1.5-hour PI; (B) pNeuO and pGapA*p*-NeuO transformed into Δ*dsdC1*, measured 1.5-hour PI; (C) pNeuO

and pGapA*p*-NeuO transformed into Δ*neuO*, measured 1.5-hour PI. Statistical non-paired Student's *t*-tests were performed using GraphPad Prism 8 between uninfected sample and infected sample, and further between infected sample and infected sample with the complement plasmid. * denotes $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$; **** $p \leq 0.0001$; ns $p > 0.05$.

Figure 5-10 (A) shows the complementation of pDsdC1 in the Δ*dsdC1* background did not restore protection against the K1 bacteriophage. Therefore, a strategy was devised in which to assemble the GapA promoter onto DsdC1 in place of its native promoter in order to facilitate a strong constitutive transcription. However, this work is still ongoing. Figure 5-10 (B) and (C) show pNeuO complemented back into Δ*dsdC1* and Δ*neuO* with its native promoter and also with the GapA promoter. pNeuO restores moderate protection from the K1 bacteriophage in the Δ*dsdC1* background (Fig. 5-10B). pGapA*p*-NeuO was able to give greater protection and in both mutant backgrounds there was significant protection from the K1 bacteriophage, compared to the infected sample (Fig. 5-10B/C). These data therefore suggests that NeuO, when sufficiently over-expressed, can protect the cell from K1 bacteriophages. Furthermore, as the pGapA*p*-NeuO complement had restored some protection from the K1 bacteriophage in the phage killing assay, the phage plaque size assay was performed to determine if the size phenotype would be restored.

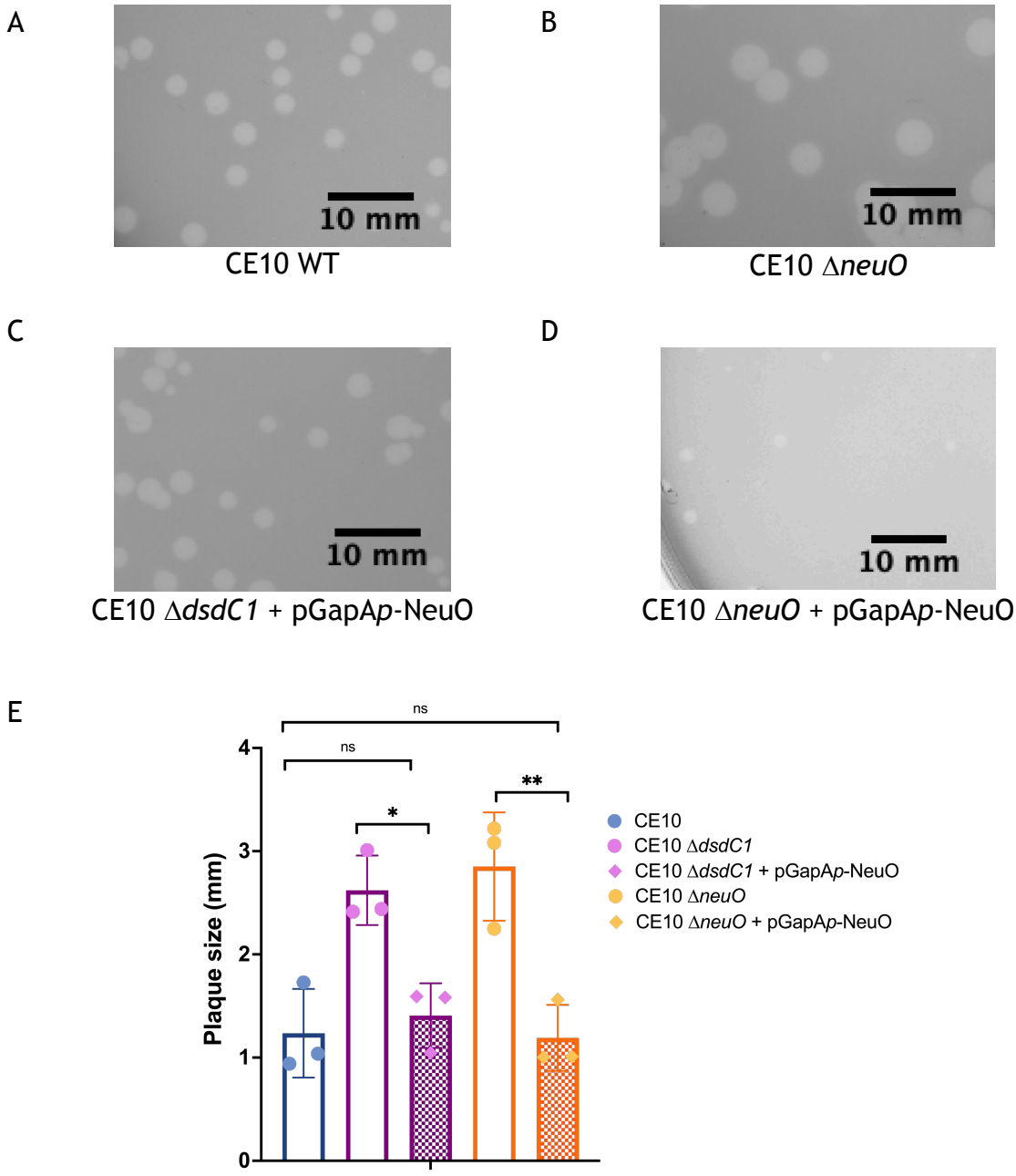


Figure 5-11 Complementation of pGapA*p*-NeuO restores the phage plaque size phenotype to WT levels. Complementation of pGapA*p*-NeuO in Δ*dsdC1* and Δ*neuO* restored the small plaque phenotype that was observed on the CE10 WT bacterial lawn. Images of phage plaques taken after infection on bacterial lawns of (A) CE10 WT; (B) Δ*neuO*; (C) Δ*dsdC1* + pGapA*p*-NeuO; (D) Δ*neuO* + pGapA*p*-NeuO. Images are representative of three independent experiments. Scale bars were generated on Fiji ImageJ software. (E) Quantitative analysis of phage plaque sizes. Statistical non-paired Student's *t*-tests were performed using GraphPad Prism 8. * denotes $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$; **** $p \leq 0.0001$; ns $p > 0.05$.

Complementing pGapA*p*-NeuO into the Δ*dsdC1* and Δ*neuO* mutants restored the small phage plaque size that was characteristic of infection on CE10 WT lawns and revealed no significant difference in plaque size between WT and complement (Fig. 5-11). Moreover, it was demonstrated that there was a significant difference in plaque size between Δ*dsdC1* and Δ*neuO* mutants, and the pGapA*p*-NeuO complemented mutants (Fig. 5-11E). Furthermore, the complementation also removed the halo effect that had been seen on the Δ*dsdC1* and Δ*neuO* bacterial lawns. When all of these data are combined, a DsdC1-specific regulatory role over *neuO* starts to emerge. It appears as though NeuO, an *O*-acetyltransferase, regulated through DsdC1, is able to mediate protection against K1 bacteriophages.

5.5.1 Phase variation of *neuO* influences susceptibility to K1 bacteriophage killing

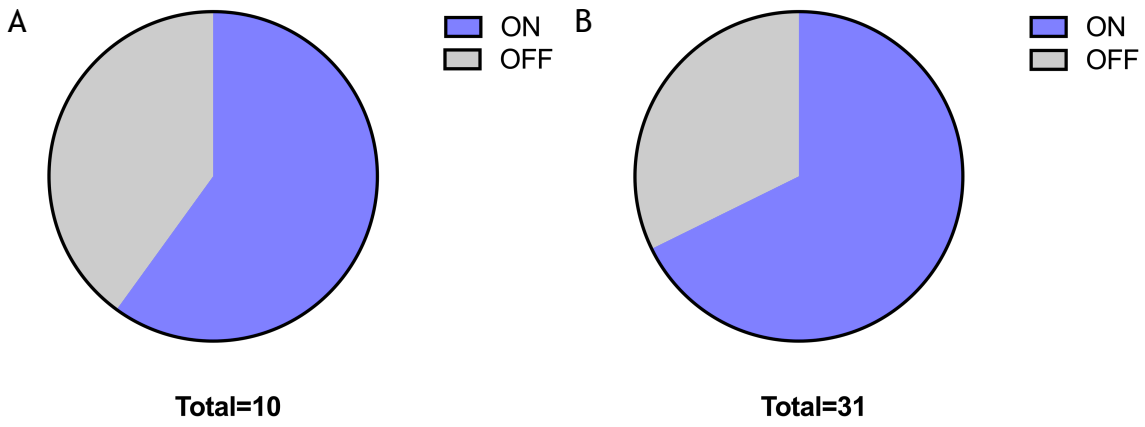
As noted above during the K1 phage killing assays, two distinct phenotypes emerged from clonal CE10 WT populations: those that were protected against the K1 bacteriophage and those that were lysed in similar rates to the Δ*dsdC1* and Δ*neuO* mutants (Fig. 5-9A). Those colonies were sequenced and revealed those that were able to protect from the K1 bacteriophage-mediated lysis were NeuO “phase ON” and those that were lysed were NeuO “phase OFF” (Table 5-1).

Sample	First 64 codons of NeuO	Phase state?	Number of 5'-AAGACTC-3' repeats
NeuO reference	MSRLKTQDSRLKTQDSRLKTQDSRLKTQ DSRLKTQDSRLKTQDSFSIDDNGSGNIF VCGDLVNS...	Phase ON	18
WT_1	MSRLKTQDSRLKTQDSRLKTQDSRLKTQ DSRLKTQDSRLKTQDSRLVFH***WVR* YFCMWRSCK*	Phase OFF	19
WT_2	MSRLKTQDSRLKTQDSRLKTQDSRLKTQ DSRLKTQDSRLKTQDSFSIDDNGSGNIF VCGDLVNS…	Phase ON	18
WT_3	MSRLKTQDSRLKTQDSRLKTQDSRLKTQ DSRLKTQDSRLKTQDSRLVFH***WVR* YFCMWEIL*I	Phase OFF	19
WT_4	MSRLKTQDSRLKTQDSRLKTQDSRLKTQ DSRLKTQDSRLKTQDSFSIDDNGSGNIF VCGDLVNS…	Phase ON	18
WT_5	MSRLKTQDSRLKTQDSRLKTQDSRLKTQ DSRLKTQDSRLKTQDSFSIDDNGSGNIF VCGDLVNS…	Phase ON	18
WT_6	MSRLKTQDSRLKTQDSRLKTQDSRLKTQ DSRLKTQDSRLKTRFPLMIMGQVIFLYV EIL*IAKR	Phase OFF	17
WT_7	MSRLKTQDSRLKTQDSRLKTQDSRLKTQ DSRLKTQDSRLKTQDSFSIDDNGSGNIF VCGDLVNS...	Phase ON	18
WT_8	MSRLKTQDSRLKTQDSRLKTQDSRLKTQ DSRLKTQDSRLKTQDSFSIDDNGSGNIF VCGDLVNS...	Phase ON	18
WT_9	MSRLKTQDSRLKTQDSRLKTQDSRLKTQ DSRLKTQDSRLKTQDSFSIDDNGSGNIF VCGDLVNS...	Phase ON	18
WT_10	MSRLKTQDSRLKTQDSRLKTQDSRLKTQ DSRLKTQDSRLKTRFPLMIMGQVIFLYV EIL*IAKR	Phase OFF	17

Table 5-1 Sequencing results of the NeuO polyΨ region in 10 CE10 WT colonies. Sequencing results of the first 192 nucleotides (64 codons) of the NeuO polyΨ region in 10 independent CE10 WT colonies. Six frame translation from the ATG start codon revealed the amino acid sequence and consequently the "phase ON/OFF" status of the colony. The red stars indicate a stop codon. The blue and green "SRLKTQD" amino acids indicate the polyΨ region. The number of 5'-AAGACTC-3' heptanucleotides in each polyΨ region of NeuO is also indicated.

Table 5-1 revealed the differences in sequence variability of the NeuO polyΨ region of the CE10 WT colonies. Indeed, although any multiple of 3 heptanucleotides would result in a "phase ON" colony, all "phase ON" colonies were 18 heptanucleotides long. Further, again, although any loss or gain of other than 3 heptanucleotides would result in a "phase OFF" colony, all the "phase

OFF" colonies were either 17 or 19 heptanucleotides long. As phase variation is described as a random event (Henderson et al., 1999); the percentage of CE10 WT colonies that were "phase ON" and "phase OFF" were measured (Fig. 5-12).



Sequencing data Experimental data

Figure 5-12 Graphical representation of the number of CE10 WT NeuO "phase ON" and "phase OFF" colonies. Number of NeuO "phase ON" and "phase OFF" colonies from (A) sequencing data and (B) experimental data from the K1 bacteriophage killing assay.

The number of "phase ON" colonies identified appeared to be greater than the number of "phase OFF" colonies (Fig. 5-12). In the experimental data, a colony was deemed "phase OFF' if it had been completely lysed or mostly lysed by hour 6 (3 hours PI). Populations that were protected from the K1 bacteriophage-mediated lysis were deemed "phase ON". From 31 K1 phage killing experiments, 68% were "phase ON" (21/31), whereas "phase OFF" colonies accounted for 32%. The sequencing data revealed a similar proportion of "phase OFF" colonies (40% (4/10)), whereas 60% (6/10) where "phase ON".

5.5.2 Phase variation affects transcriptional levels of *neuO*

An RT-qPCR comparison of *neuO* in the Δ*dsdC1* mutant background, in relation to the CE10 WT, showed a 4.3-fold reduction in expression (Fig. 5-13A). However, the *p*-value was not significant at 0.051. As the phase state of NeuO had

affected the CE10 WT phage killing assay, it was hypothesised that it may also have an effect on observable transcript levels. For this experiment, colonies were deemed “phase ON” or “phase OFF” due to prior survival or lysis in the K1 bacteriophage killing assay.

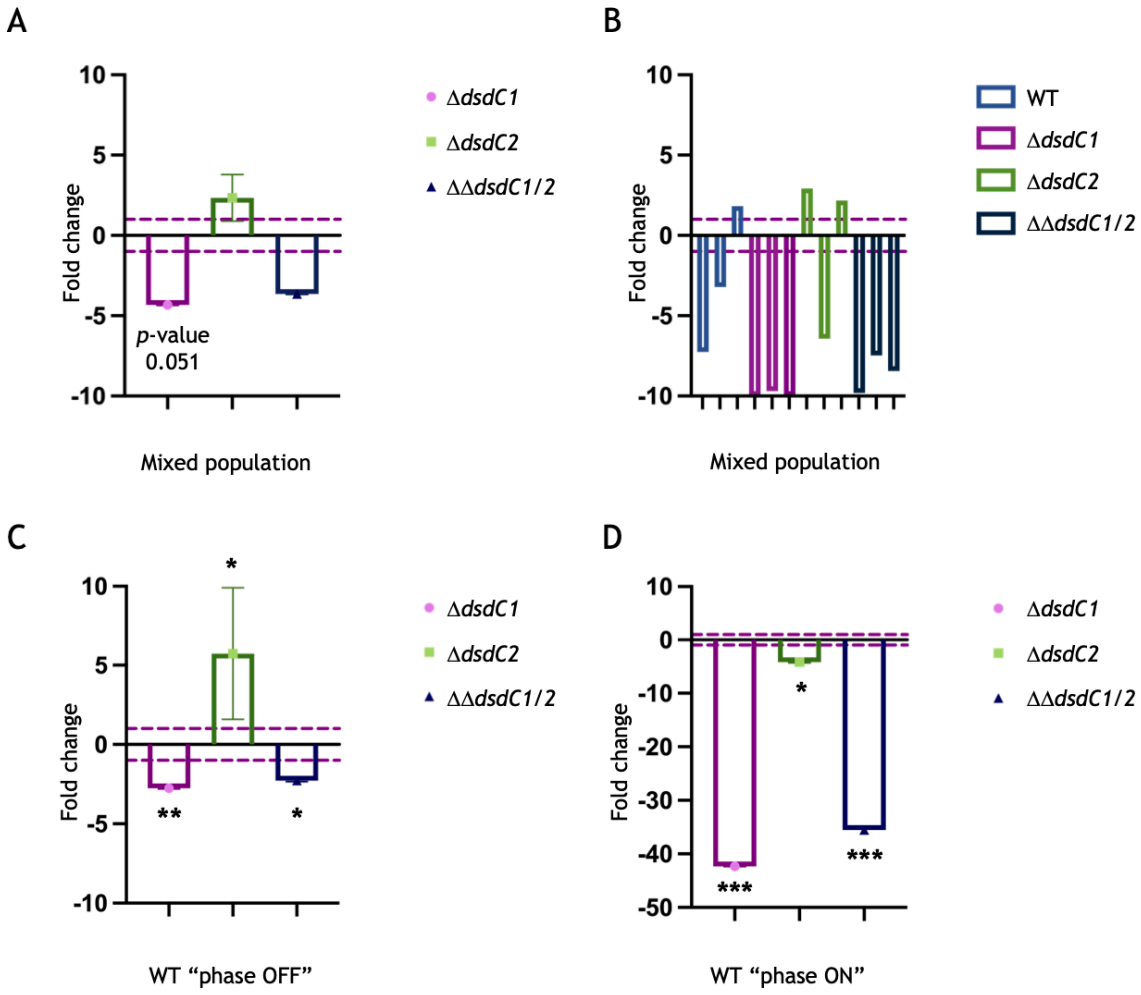


Figure 5-13 Phase variation of NeuO affects transcriptional expression. Phase variation affects the transcription of *neuO*. (A) The WT was a mixed population of “phase ON” and “phase OFF” colonies. (B) Individual expression values for the colonies used in (A). (C) The WT was a population of “phase OFF” colonies. (D) The WT was a population of “phase ON” colonies. The purple dashed line indicates baseline expression of *neuO* in CE10 WT. Gene expression analyses were based off three biological repeats. * denotes $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$; **** $p \leq 0.0001$; ns $p > 0.05$.

Figure 5-13 indicated that the phase variation of NeuO in the CE10 WT affected transcriptional expression. Indeed, when the WT was a mixed population, there was no significant difference in fold changes in the Δ*dsdC1* and ΔΔ*dsdC1/2*

backgrounds, -4.32 (*p*-value 0.051) and -3.64 (*p*-value 0.1) respectively (Fig. 5-13A). However, when three "phase ON" colonies were used, there was a significant reduction in expression of *neuO* -42.35-fold (*p*-value 0.0004) in the Δ*dsdC1* background and -35.51-fold (*p*-value 0.0004) in the ΔΔ*dsdC1/2* background (Fig. 5-13D). Moreover, when three WT "phase OFF" colonies were used as the control, *neuO* expression was reduced to -2.75-fold (*p*-value 0.006) and -2.28-fold (*p*-value 0.01) in the Δ*dsdC1* and ΔΔ*dsdC1/2* backgrounds respectively (Fig. 5-13C). Although these colonies were not sequenced, it is clear that the WT and Δ*dsdC2* colonies have variable levels of *neuO* transcript, whereas the Δ*dsdC1* and ΔΔ*dsdC1/2* colonies did not (Fig 5-13B). Here it has been shown that the WT and Δ*dsdC2* display both high and low levels of *neuO* transcript, whereas Δ*dsdC1* was consistently low. This is consistent with the previous findings of variation in the CE10 and Δ*dsdC2* backgrounds, in the K1 bacteriophage killing assay, compared to the relative consistency of the Δ*dsdC1* and ΔΔ*dsdC1/2* backgrounds that was observed. Due to the consistency observed from Δ*dsdC1* in the K1 bacteriophage assay and consistently low levels of *neuO* transcript, it was hypothesised that potentially the extent of NeuO phase switching was playing a role in the Δ*dsdC1* phenotype. To test this, 10 Δ*dsdC1* and 10 Δ*dsdC2* colonies were sequenced at the NeuO polyΨ region.

Δ*dsdC1*					Δ*dsdC2*			
Sample	**First 64 codons of NeuO**	**Phase state?**	**Number of 5'-AAGACTC-3' repeats**		**Sample**	**First 64 codons of NeuO**	**Phase state?**	**Number of 5'-AAGACTC-3' repeats**
NeuO reference	MSRLKTQDSRLKTQDSRLKTQDSRLKTQDSRLKTQDSRLKTQDSFSIDDNGSGNIFVCGDLVNS...	Phase ON	18		NeuO reference	MSRLKTQDSRLKTQDSRLKTQDSRLKTQDSRLKTQDSRLKTQDSFSIDDNGSGNIFVCGDLVNS...	Phase ON	18
Δ*dsdC1*_1	MSRLKTQDSRLKTQDSRLKTQDSRLKTRFPLMIMGQVIFLYVEIL*IAKRIKFSSMETITNLL*	Phase OFF	11		Δ*dsdC2*_1	MSRLKTQDSRLKTQDSRLKTQDSRLKTQDSRLKTQDSRLKTQDSFSIDDNGSGNIFVCGDLVNS…	Phase ON	18
Δ*dsdC1*_2	MSRLKTQDSRLKTQDSRLKTQDSRLKTRFPLMIMGQVIFLYVEIL*IAKRIKFSSMETITNLL*	Phase OFF	11		Δ*dsdC2*_2	MSRLKTQDSRLKTQDSRLKTQDSRLKTQDSRLKTQDSRLKTQDSFSIDDNGSGNIFVCGDLVNS…	Phase ON	18
Δ*dsdC1*_3	MSRLKTQDSRLKTQDSRLKTQDSRLKTRFPLMIMGQVIFLYVEIL*IAKRIKFSSMETITNLL*	Phase OFF	11		Δ*dsdC2*_3	MSRLKTQDSRLKTQDSRLKTQDSRLKTQDSRLKTQDSRLKTQDSFSIDDNGSGNIFVCGDLVNS…	Phase ON	18
Δ*dsdC1*_4	MSRLKTQDSRLKTQDSRLKTQDSRLKTRFPLMIMGQVIFLYVEIL*IAKRIKFSSMETITNLL*	Phase OFF	11		Δ*dsdC2*_4	MSRLKTQDSRLKTQDSRLKTQDSRLKTQDSRLKTQDSRLKTQDSFSIDDNGSGNIFVCGDLVNS…	Phase ON	18
Δ*dsdC1*_5	MSRLKTQDSRLKTQDSRLKTQDSRLKTRFPLMIMGQVIFLYVEIL*IAKRIKFSSMETITNLL*	Phase OFF	11		Δ*dsdC2*_5	MSRLKTQDSRLKTQDSRLKTQDSRLKTQDSRLKTQDSRLKTQDSFSIDDNGSGNIFVCGDLVNS…	Phase ON	18
Δ*dsdC1*_6	MSRLKTQDSRLKTQDSRLKTQDSRLKTRFPLMIMGQVIFLYVEIL*IAKRIKFSSMETITNLL*	Phase OFF	11		Δ*dsdC2*_6	MSRLKTQDSRLKTQDSRLKTQDSRLKTQDSRLKTQDSRLKTQDSFSIDDNGSGNIFVCGDLVNS…	Phase ON	18
Δ*dsdC1*_7	MSRLKTQDSRLKTQDSRLKTQDSRLKTRFPLMIMGQVIFLYVEIL*IAKRIKFSSMETITNLL*	Phase OFF	11		Δ*dsdC2*_7	MSRLKTQDSRLKTQDSRLKTQDSRLKTQDSRLKTQDSRLKTQDSFSIDDNGSGNIFVCGDLVNS…	Phase ON	18
Δ*dsdC1*_8	MSRLKTQDSRLKTQDSRLKTQDSRLKTRFPLMIMGQVIFLYVEIL*IAKRIKFSSMETITNLL*	Phase OFF	11		Δ*dsdC2*_8	MSRLKTQDSRLKTQDSRLKTQDSRLKTQDSRLKTQDSRLKTQDSFSIDDNGSGNIFVCGDLVNS…	Phase ON	18
Δ*dsdC1*_9	MSRLKTQDSRLKTQDSRLKTQDSRLKTRFPLMIMGQVIFLYVEIL*IAKRIKFSSMETITNLL*	Phase OFF	11		Δ*dsdC2*_9	MSRLKTQDSRLKTQDSRLKTQDSRLKTQDSRLKTQDSRLKTQDSFSIDDNGSGNIFVCGDLVNS…	Phase ON	18
Δ*dsdC1*_10	MSRLKTQDSRLKTQDSRLKTQDSRLKTRFPLMIMGQVIFLYVEIL*IAKRIKFSSMETITNLL*	Phase OFF	11		Δ*dsdC2*_10	MSRLKTQDSRLKTQDSRLKTQDSRLKTQDSRLKTQDSRLKTQDSFSIDDNGSGNIFVCGDLVNS…	Phase ON	18

Table 5-2 Sequencing results of the NeuO polyΨ region in 10 CE10 Δ*dsdC1* and Δ*dsdC2* colonies. Sequencing results of the first 192 nucleotides (64 codons) of the NeuO polyΨ region in 10 independent CE10 Δ*dsdC1* and Δ*dsdC2* colonies. Six frame translation from the ATG start codon revealed the amino acid sequence and consequently the "phase ON/OFF" status of the colony. The red stars indicate a stop codon. The number of 5'-AAGACTC-3' heptanucleotides in each polyΨ region of NeuO is also indicated.

The data contained within Table 5-2 revealed that all of the Δ*dsdC1* mutants that were sequenced were "phase OFF" whereas all of the Δ*dsdC2* mutants were "phase ON". Intriguingly, all of the Δ*dsdC1* "phase OFF" colonies were 11 heptanucleotides long, whereas the Δ*dsdC2* "phase ON" colonies were all 18 heptanucleotides long. The proportion of experimentally determined Δ*dsdC1* and Δ*dsdC2* NeuO "phase ON" and "phase OFF" colonies from the K1 bacteriophage killing assay were also measured (Fig. 5-14).

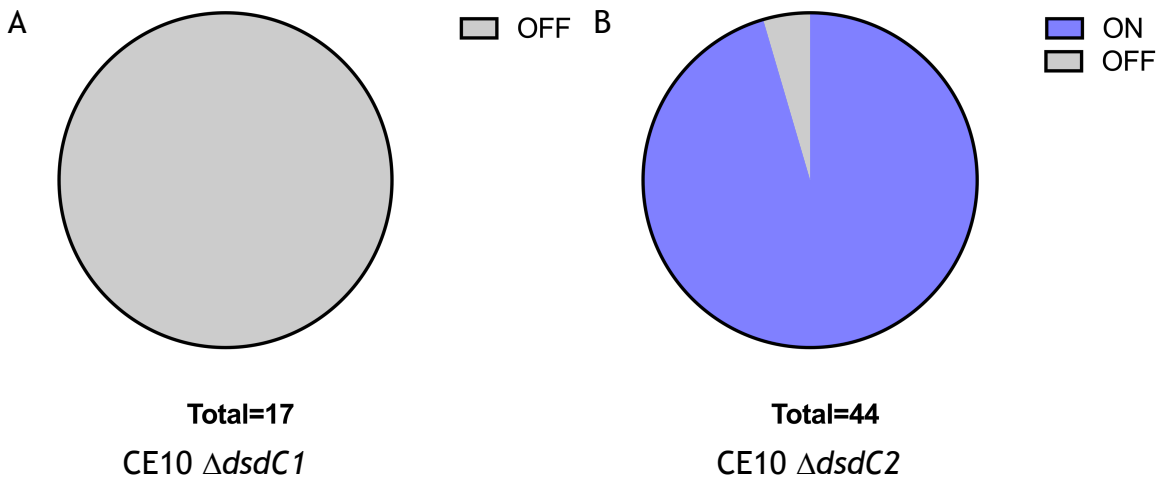


Figure 5-14 Δ*dsdC1* is constituently NeuO "phase OFF". Number of NeuO "phase ON" and "phase OFF" colonies from the K1 bacteriophage killing assay experimental data for (A) Δ*dsdC1* and (B) Δ*dsdC2*.

Using the K1 bacteriophage killing assay data, "phase ON" and "phase OFF" states were assigned to the Δ*dsdC1* and Δ*dsdC2* mutants. A colony was deemed "phase OFF" if it had been completely lysed or mostly lysed by hour 6 (3 hours PI). Those that were still able to mediate protection from the K1 bacteriophage were deemed "phase ON". Figure 5-14 revealed all of the Δ*dsdC1* colonies were NeuO "phase OFF". By hour 6, the cells had been completely lysed in 100% of experiments (17/17). This corresponds with the sequencing data, where 100% of the colonies sequenced were "phase OFF". It appears therefore that Δ*dsdC1* has a permanently "phase OFF" NeuO. Conversely, for the Δ*dsdC2* mutant, although it appeared in the sequencing data that all of the colonies were "phase ON", in the phage killing assay there were two occasions out of 44 were Δ*dsdC2* colonies

were “phase OFF” (4.5%). Although this is significantly lower than the WT “phase OFF” variability of ~32%, it does imply that the Δ*dsdC2* mutant may also be able to undergo phase switching, but at a lower rate. Taken together, these data suggests that DsdC1 directly regulates the phase state of *neuO,* however the underlying mechanism of this regulation was not investigated and will be subject to further work.

5.6 Discussion

5.6.1 The CUS-3 prophage is not present in the CE10 genome

When *neuO* was first described in 2005, it was suggested to be genetically linked to the prophage CUS-3 (Deszo *et al.*, 2005). CUS-3 was integrated into the RS218 genome in the hypervariable region between *dsdC* and *argW*. Indeed in the results reported, the strains lacking CUS-3, lacked *O*-acetyltransferase activity (Deszo *et al.*, 2005), thereby suggesting without CUS-3, *neuO* was not present on the genome. Further studies revealed that CUS-3 was an active mobile contingency locus in K1 *E. coli,* able to infect sensitive K1 hosts (King *et al.*, 2007); thereby disseminating *neuO* further into the population. However, a later study of 111 K1 clinical isolates revealed that although 61% carried the *neuO* gene, 11% of those NeuO-positive strains lacked CUS-3 (King *et al.*, 2007). This indicated for the first time that CUS-3 has been potentially excised or lost from these strains, leaving *neuO*, or that *neuO* was potentially integrated into the genome on a different MGE. PHASTER analysis of the CE10 genome revealed 15 different genomic regions in which prophages had been identified, including an incomplete prophage of 16.2 kb between nucleotide positions 824561-840854 bp, immediately upstream of the endosialidase gene and *neuO*. PHASTER analysis revealed the phage with the closest homology was the HK620 phage. HK620 phages encode a tail-spike protein and can recognise host cell receptor polysaccharides in *E. coli* (Barbirz *et al.*, 2008). Intriguingly, CUS-3 has been reported to be a close relative of HK620, differing in CUS-3’s acquisition of *neuO* and the endosialidase gene (King *et al.*, 2007).

Further BLAST analysis, using EasyFig, revealed homology between the incomplete prophage and CUS-3 at the start and at the end of the prophage,

with 37 genes in the centre having no homology (Fig. 5-3). As the PHASTER analysis revealed, the incomplete prophage is only 16.2 kb long, which indicates that it is not big enough to be a different prophage to CUS-3 and is likely therefore to be either a degraded CUS-3, or *neuO* is encoded on a separate MGE (Touchon *et al.*, 2016). Due to the high homology between the two prophage regions, this potentially indicates that *neuO* was likely integrated into the CE10 genome on CUS-3, however CUS-3 has subsequently been lost. Indeed, over time, prophage genes can be lost from the genome and prophages can be degraded so much that the few genes remaining can no longer be characterised as a prophage (Touchon *et al.*, 2016). Analysis of bacterial genomes revealed that after successful lysogenisation into the bacterial host, prophages undergo a complex decay, consisting of genomic rearrangements, inactivation of genes due to point mutations, and modular exchanges (Canchaya *et al.*, 2003). This "domestication" of prophages allows the bacteria to retain desired functions, which in the case of *neuO* is the ability to *O*-acetylate the capsule, whilst removing a molecular "time bomb" (Bobay *et al.*, 2014). Further, analysis has indicated that genes encoded in the central part of the prophage are more frequently deleted than genes at the edges (Bobay *et al.*, 2014), which appears to be the case in the incomplete prophage in CE10 and further indicates that CUS-3 originally integrated into the CE10 genome, but has since been lost. However, *neuO* has been retained, suggesting its importance to CE10.

5.6.2 DsdC regulates *neuO*, the *O*-acetyltransferase

Using the ChIP-Seq analysis it was shown that DsdC, the D-serine metabolism locus regulator, binds to the DNA sequence encoding NeuO, the *O*-acetyltransferase. Interestingly the binding peak of DsdC1 and DsdC2 in *neuO* was in the ORF of the gene. As discussed in section 3.6.2, intragenic binding peaks have become more common in the analysis of ChIP-Seq data sets and this could be for a multitude of reasons (Galagan *et al.*, 2012). Often however it is only repressors that bind after the TSS, with one study showing the binding sites of activators only occurred before the TSS (Babu & Teichmann, 2003b). This suggests therefore that DsdC is repressing *neuO*. However, directly testing this hypothesis using transcriptomics and RT-qPCR data contradicts this assumption, as in the Δ*dsdC1/2* mutant background *neuO* is down-regulated, indicating that DsdC is an activator of *neuO*.

Transcription is a tightly controlled process within bacteria and is often regulated in response to environmental signals and cues. Investigating the regulation of *neuO* by DsdC in the presence and absence of D-serine, revealed new insights into the role of this metabolite. The RNA-Seq data presented in Chapter 3 revealed that in the absence of D-serine, *neuO* expression was reduced -7.9-fold in the ΔΔ*dsdC1/2* mutant background. Upon exposure to D-serine, the fold-change increased to -4.57-fold, indicating that when D-serine was present, DsdC reduces its activation of *neuO*. However, when pDsdA was transformed into ΔΔ*dsdC1/2*, *neuO* expression decreased -8.15-fold, similar to the *neuO* levels of transcript observed in the absence of D-serine data. This suggests that the accumulation of D-serine affected the transcription of *neuO*, rather than exposure to D-serine. Overall, this indicated that DsdC regulates *neuO* independently of D-serine presence.

As discussed above, *O*-acetylation of the capsule can change the biophysical properties of the capsular polysaccharides (Fusco *et al.*, 2007), as well as potentially affecting the virulence status of the bacterial cell (Frasa *et al.*, 1993). Intriguingly, non-enzymatic acetylation has been suggested to be in response to environmental signals and the metabolism status of the cell (reviewed in Christensen *et al.*, 2019). Indeed, depending on the concentrations of glucose, lactate, fructose, and xylose, these carbon sources have all been reported to potentially affect acetylation in *E. coli* (reviewed in Christensen *et al.*, 2019). As DsdC is the D-serine metabolism locus regulator, this could indicate a putative link between *O*-acetylation of the capsule and the host environment.

5.6.3 NeuO mediates protection from K1 bacteriophage lysis in CE10

Prophages that encode receptor modifying enzymes often do so to prevent further phage entry into the host bacterial cell, thereby stopping bacteriophage "super-infections" (reviewed in Labrie, Samson & Moineau, 2010). Indeed, a recent study reported that 29 out of 30 prophages that could infect *Pseudomonas aeruginosa* mediated resistance to at least 3 phages (Bondy-Denomy *et al.*, 2016). Furthermore, 12 of these phages had the capacity to mediate resistance to over 20 different phages (Bondy-Denomy *et al.*, 2016).

From the CE10 genome sequence, it indicated that *neuO* appeared to have been integrated into the genome by a MGE (Fig. 5-3). To test the hypothesis that NeuO-mediated *O*-acetylation modifies the bacteriophage receptor on the capsule surface, and this in turn is regulated by DsdC, a K1 bacteriophage titre and killing assays were performed on CE10 and isogenic mutants.

The common assumption is that each plaque present on a plate is initiated by a single virus particle (Gallet *et al.*, 2011). As can be seen in Figure 5-5, there was no significant difference between the ability of the K1 bacteriophage to form plaques on CE10 and isogenic mutants. Therefore, it was assumed that the K1 bacteriophage is able to infect CE10 and the mutants at similar frequency. Strikingly however, there was a difference between WT and some of the mutants in plaque size and morphology (Fig. 5-6). Circular plaque morphology is due to a cycle of infections of the bacterial cell by bacteriophages disseminating outwards from the original infected host (Gallet *et al.*, 2011). All of the plaques formed by the K1 phage were clear and had a clean periphery, typical for lytic virulent phages (Jurczak-Kurek *et al.*, 2016). However, some of the plaques formed on Δ*dsdC1*, ΔΔ*dsdC1/2* and Δ*neuO* had haloes surrounding them (Fig. 5-7). Haloes have been described as semi-transparent zones around the plaque, which have been suggested to be formed by phage-produced depolymerases, which are destroying the bacterial cell surface structures (Jurczak-Kurek *et al.*, 2016; Pires *et al.*, 2016). Bacteriophages produce degrading enzymes, for example sialidases, in order to cleave the glycosidic bonds in the capsule, enabling adsorption and invasion into the host bacteria (Pires *et al.*, 2016). Intriguingly, *O*-acetylation of bacterial cell wall polysaccharides have been shown to increase resistance to degradative enzymes (Bernard *et al.*, 2011). As the haloes indicated an increase in depolymerase activity of the K1 bacteriophage on the Δ*dsdC1*, ΔΔ*dsdC1/2* and Δ*neuO* bacterial lawns, this indicated that potentially these mutants were unable to *O*-acetylate their capsule, thus allowing for increased depolymerisation.

Furthermore, there was also a difference in the size of the plaques formed. Plaques formed on Δ*dsdC1*, ΔΔ*dsdC1/2* and Δ*neuO* were significantly larger than plaques formed on CE10 WT and Δ*dsdC2* (Fig. 5-6). A recent paper classed plaques that were ≤1 mm in size as 'small' and plaques that were ≥2 mm as

'large' (Montso *et al.*, 2019). Clearly, the data presented here shows that the plaques formed on the WT and Δ*dsdC2* were small-medium, whereas the plaques formed on Δ*dsdC1*, ΔΔ*dsdC1/2* and Δ*neuO* were large plaques. As phage replication often ceases when the bacteria reach stationary phase (Abedon, 2021), the size of a plaque can reveal several interesting characteristics of the behaviour of the bacteriophage. Although larger plaques could be due to a number of reasons including, shorter phage latent periods, virion physical properties and greater nutrient availability (Abedon, 2021), it was assumed due to the conditions staying constant between all of the experiments, that the only difference was the mutations in the bacterial hosts, and corresponding changes in the capsule. Therefore, the size of plaque likely corresponded to how quickly the virion diffused out of the initial infected cell and subsequent bacterial lysis. K1 bacteriophages use tailspike proteins, that act as capsule depolymerases, to recognise and bind to the host bacterial capsular polysaccharide (Stummeyer *et al.*, 2006). *O*-acetylation of bacterial outer surface structures has been shown to effect phage adsorption, with acetyl groups on the O-antigen protecting *E. coli* 4s from infection (reviewed in Prokhorov *et al.*, 2017). Furthermore, some bacteriophages have counter-evolved enzymes that can deacetylate the cell surface structures, thereby allowing infection (Prokhorov *et al.*, 2017). The bacteriophage G7C encodes the tailspike protein gp63.1, which can deacetylate the polysaccharide of *E. coli* 4s, whilst leaving the polysaccharide backbone intact (Prokhorov *et al.*, 2017), thus enabling recognition and subsequent infection of the cell. These results therefore potentially indicate that the *O*-acetylation of the K1 capsule, mediated by the *O*-acetyltransferase NeuO and activated by DsdC1, inhibits K1 bacteriophage dissemination in CE10, evidenced by the smaller plaque sizes observed in the WT and Δ*dsdC2* mutant.

Furthermore, using a K1 bacteriophage killing assay, it was shown that Δ*neuO* and Δ*dsdC1* were "hyper-susceptible" to the K1 bacteriophage and almost entirely lysed 1.5-hours PI, whereas the WT and Δ*dsdC2* were not lysed until 3-hours PI (Fig. 5-9). This again further supports the hypothesis that NeuO, regulated by DsdC1, modifies the K1 capsule by *O*-acetylation of the polysialic acid, thereby preventing immediate lysis by the K1 bacteriophage. Furthermore, overexpression of NeuO partially restored the survival phenotype observed of the WT (Fig. 5-10) and fully restored the plaque size phenotype observed on the WT

bacterial lawn (Fig. 5-11). These data combined suggests that *neuO* regulation is a DsdC1-mediated specific phenotype, not facilitated by DsdC2, showing different roles for these highly homologous TFs.

Strikingly, original work by Vimr and colleagues suggested that NeuO did not mediate protection against K1 bacteriophages (King, Steenbergen & Vimr., 2007). However, the data presented here clearly supports the hypothesis that NeuO mediates protection against K1 bacteriophages. The differences in data reported could be due to the differences in genetic makeup between the two strains. Vimr and colleagues used the strain RS218, which has the full CUS-3 prophage including *neuO* encoded on its genome, unlike CE10 which has a degraded incomplete prophage. Further, *neuO* is integrated downstream of *dsdC* in RS218, whereas in CE10, *neuO* is integrated 2 mb away from *dsdC*. These differences in genetic makeup may have led to the observation that NeuO does not protect against K1 bacteriophage, and further highlights how even highly conserved genes may play differing roles in strains of the same pathotype.

NeuO was also shown to be responsible for protection against K1 bacteriophage lysis, due to the two different WT phenotypes that arose from the phage killing assay. It was established through sequencing that the WT colonies that were able to protect against K1 bacteriophage were “phase ON”, whereas the WT colonies that were not able to protect against K1 bacteriophage were “phase OFF”. Phase variation of the capsule is not uncommon, indeed as stated above, it is a mechanism used by multiple bacteria and fungi. Why though, if *O*-acetylation of the capsule prevents K1 bacteriophage-mediated lysis, is the capsule not *O*-acetylated all of the time? Certainly, it has been reported that the degree of acetylation can vary between isolates, ranging from 5-95% (King, Steenbergen & Vimr., 2007). However, by modifying capsular states by phase variation, this allows bacteria to evade specific immune responses against one capsule form. The K1 polysialic capsule mimics host N-CAM surfaces, thus minimising immune responses from the host (Silver & Vimr, 1990). However, further heterogeneity of the population, mediated through *O*-acetylation of the capsule, means that in an event of antibody recognition to one phase state, the other will survive and proliferate (Van Der Woude & Bäumler, 2004). Further, recent research has suggested that *O*-acetylation plays a role in environmental survival, with *O*-acetylated strains more resistant to desiccation, but having a

reduced ability to form biofilms (Mordhorst *et al.*, 2009). Therefore, like most phase variable genes there is a delicate balance between phase switching, likely in relation to the host environment.

5.6.4 Phase variation as a putative indirect mechanism of transcriptional regulation of *neuO*

The data above suggests that a DsdC1-specific regulatory mechanism mediates transcription of *neuO*. Interestingly however, the RT-qPCR data suggested that phase state of the colonies correlated with the level of transcription of *neuO* (Fig. 5-13). Indeed, when the WT population was a mix between "phase ON" and "phase OFF" colonies, there was a -4.32 decrease in fold change of *neuO* between WT and Δ*dsdC1*. However, when the WT population was selected for only "phase ON" colonies, there was a decrease in fold change of -42.35 between WT and Δ*dsdC1*. Moreover, when the WT population was comprised of "phase OFF" colonies the fold change was -2.75. Although these colonies were not sequenced, they were determined to be "phase ON" or "phase OFF" using the K1 bacteriophage killing assay.

This suggests therefore that phase variation affects transcription of *neuO*. In CE10, it is the random event of SSM that causes NeuO to be "phase ON" or "phase OFF". SSM commonly occurs in genes that benefit from phase variation between "ON" and "OFF" states, such as those involved the production of antigenic surface appendages. Indeed, SSM occurs in the *Neisseria spp*. for phase variation of the capsule (reviewed in Tzeng, Thomas & Stephens, 2016). SSM can affect transcription or translation, dependent on where the repeated homologous regions are (Henderson *et al.*, 1999). The SSM of *neuO* occurs within the ORF of the gene, and hence, alters the reading frame of the mRNA produced. SSM also causes phase variation of the *N. meninigitis* capsule, mediated via the polysialyltransferase gene *siaD* (Hammerschmidt *et al.*, 1996). Interestingly, it has been reported that transcription of the "phase OFF" *siaD* can be prematurely terminated (Lavitola *et al.*, 1999). It was shown that Rho, the canonical transcription termination factor, was responsible for premature transcription termination of the out of frame *siaD* (Lavitola *et al.*, 1999). Potentially therefore, a similar mechanism is occurring in CE10, and the "phase OFF" *neuO* is prematurely transcriptionally terminated, thereby explaining how

phase variation is affecting the transcriptional levels in the RT-qPCR. Interestingly it appears from this data that DsdC1 affects the SSM of NeuO. When Δ*dsdC1* is knocked out, this caused *neuO* to remain in a "locked off" position (Table 5-2).

5.6.5 Conclusion

It has been stated that up to 50% of prokaryotic genes may have arisen from gene duplication (Brenner *et al.*, 1995; Teichmann *et al.*, 1998; Teichmann & Babu, 2004). In 2006, it was noted that in K1 strains, there were two genetically unlinked copies of the *dsdCXA* locus, with both copies retaining functionality for metabolising D-serine (Moritz & Welch, 2006). When TFs evolve by duplication, regulatory networks are either inherited or lost (Teichmann & Babu, 2004). Although Chapter 3 indicated that, in CE10, DsdC1 and DsdC2 regulated the same set of genes, the results presented in Chapter 5 have revealed novel differing roles between the highly homologous proteins. Indeed, the data in Chapter 5 has revealed novel phenotypical properties of DsdC1 in *E. coli* K1 pathogenicity. The data presented here have shown that DsdC1 mediates protection from K1 bacteriophage lysis, through the regulation of *neuO,* likely therefore affecting the *O*-acetylation state of the capsule and thus blocking K1 bacteriophage receptors. This work indicates that DsdC1 has been tailored in CE10 in a strain-specific manner.

Chapter 6 Investigating further putative roles of DsdC in uropathogenic *Escherichia coli* and neonatal meningitis-associated *Escherichia coli*

6.1 Introduction

Whilst *E. coli* is an incredibly diverse species, a key virulence trait shared by all *E. coli* is the presence of the LPS (Ebbensgaard *et al.*, 2018). LPS is an integral element of the *E. coli* outer membrane, protecting the cell from the host environment and providing resistance to host complement (Heinrichs *et al.*, 1998; Lerouge & Vanderleyden, 2002). *E. coli* LPS comprises of three components: lipid A (the hydrophobic anchor that is embedded in the cell outer membrane); the core oligosaccharide (core OS) composed of sugar molecules; and the O-antigen, a structurally variable polysaccharide composed of several repeating side chains (Whitfield *et al.*, 1999). There are over 170 O-antigen serogroups that have been found in *E. coli*, reflecting the hypervariability of the polysaccharide (Whitfield *et al.*, 1999). Conversely, there is much less structural diversity within the core OS, with only 5 currently known outer core OS structures in *E. coli*; R1, R2, R3, R4, and K12 (Whitfield *et al.*, 1999). In *E. coli*, the core OS biosynthesis genes are encoded on the *waa* locus (Heinrichs *et al.*, 1998). Analysis of the CE10^{DsdC} ChIP-Seq data revealed that DsdC bound to the *waa* locus. Furthermore, upon exposure to D-serine, the transcriptomic data revealed that several of the *waa* genes were differentially regulated, indicating a direct regulatory role for DsdC in LPS biosynthesis.

Another key virulence determinant for some *E. coli* strains is the presence of an extracellular polysaccharide layer, the capsule. There are over 70 distinct capsular (K) antigens in *E. coli,* differing in several properties including sugar residues, branching and charge density (Jann & Jann, 1992). The capsule is of particular biological importance, with some K-antigens known to play roles in immune evasion and serum resistance (Jann & Jann, 1992). In *E. coli,* a ~20 kb region encodes the capsule biosynthesis genes, termed the *kps* island (Silver & Vimr, 1990). The *kps* island is organised into three adjacent genomic regions: Regions 1, 2 and 3 (Vimr *et al.*, 1995). Regions 1 and 3 are generally conserved amongst group 2 capsule expressing strains, and encode for proteins involved in the transport of the capsular polysaccharides to the cell surface, whereas Region 2 is serotype specific, and therefore in K1 strains it encodes the sialic acid biosynthesis proteins (reviewed in Whitfield & Roberts, 1999). The capsule is transcribed from two major promoter elements, PR1 and PR3, that sit upstream of Regions 1 and 3 respectively (reviewed in Jia *et al.*, 2017). Comparisons of

the CE10 ChIP-Seq and RNA-Seq data had revealed DsdC differentially expresses genes encoded on the *kps* island, thus indicating a direct regulatory role for DsdC in K1 capsule synthesis.

As well as protecting themselves from the host environment through LPS and capsule synthesis, *E. coli* can mediate pathogenesis by the production of toxins. Colibactin is a bacterial-produced genotoxin that interferes with the eukaryotic cell cycle, inducing DNA double stranded breaks (DSB) and cellular senescence (Secher *et al.*, 2013). Colibactin-producing *E. coli* are over-represented in CRC patients, with reports suggesting they are found in up to 67% of CRC patients, but in less than 20% of controls (reviewed in Faïs *et al.*, 2018). The colibactin synthesis genes are encoded on a 54-kb island, termed the *pks* island (Nougayrède *et al.*, 2006), and is mainly carried in the B2 phylogroup, which typically comprise ExPEC strains (reviewed in Faïs *et al.*, 2018). ClbR is the transcriptional activator for colibactin gene expression and is encoded within the *pks* island (Wallenstein *et al.*, 2020). Analysis of the CFT073^{DsdC} ChIP-Seq and RT-qPCR data revealed DsdC bound and differentially regulated *clbR*, thus indicating a direct role of DsdC in colibactin synthesis.

In Chapter 5, it was described how DsdC affected the regulation of *neuO*, the capsule *O*-acetyltransferase. Using a combination of the ChIP-Seq and RNA-Seq data, as indicators of potential regulation, this chapter will focus on any further phenotypes that were mediated by DsdC and the host metabolite D-serine in CE10 and CFT073; namely LPS expression in CE10, invasion and adhesion to an *in vitro* BBB cell model, and colibactin synthesis in CFT073.

6.2 DsdC modulates discrete genes involved in CE10 lipopolysaccharide expression

As discussed above, *E. coli* LPS comprises of three components: lipid A, core OS, and the O-antigen (Yethon *et al.*, 2000). Using the CE10 ChIP-Seq data, DsdC was shown to occupy several sites along the core OS biosynthesis region, the *waa* locus, binding in a broad manner across the region (Fig. 6-1A). Nucleotide analysis of the *waa* region in the UPEC strain CFT073 revealed that the region was A:T-rich, with the GC content comprising between 33-38% of the sequence (Fig. 6-1C). BLAST analysis of the *waa* region revealed 99% homology over 100%

of the sequence between CFT073 and CE10, indicating that the *waa* region in CE10 is also A:T-rich. The *waa* locus is comprised of several glycosyltransferases which sequentially elongate the core OS from a lipid A acceptor molecule, to the O-antigen attachment (Fig. 6-1B) (Heinrichs *et al.*, 1998).

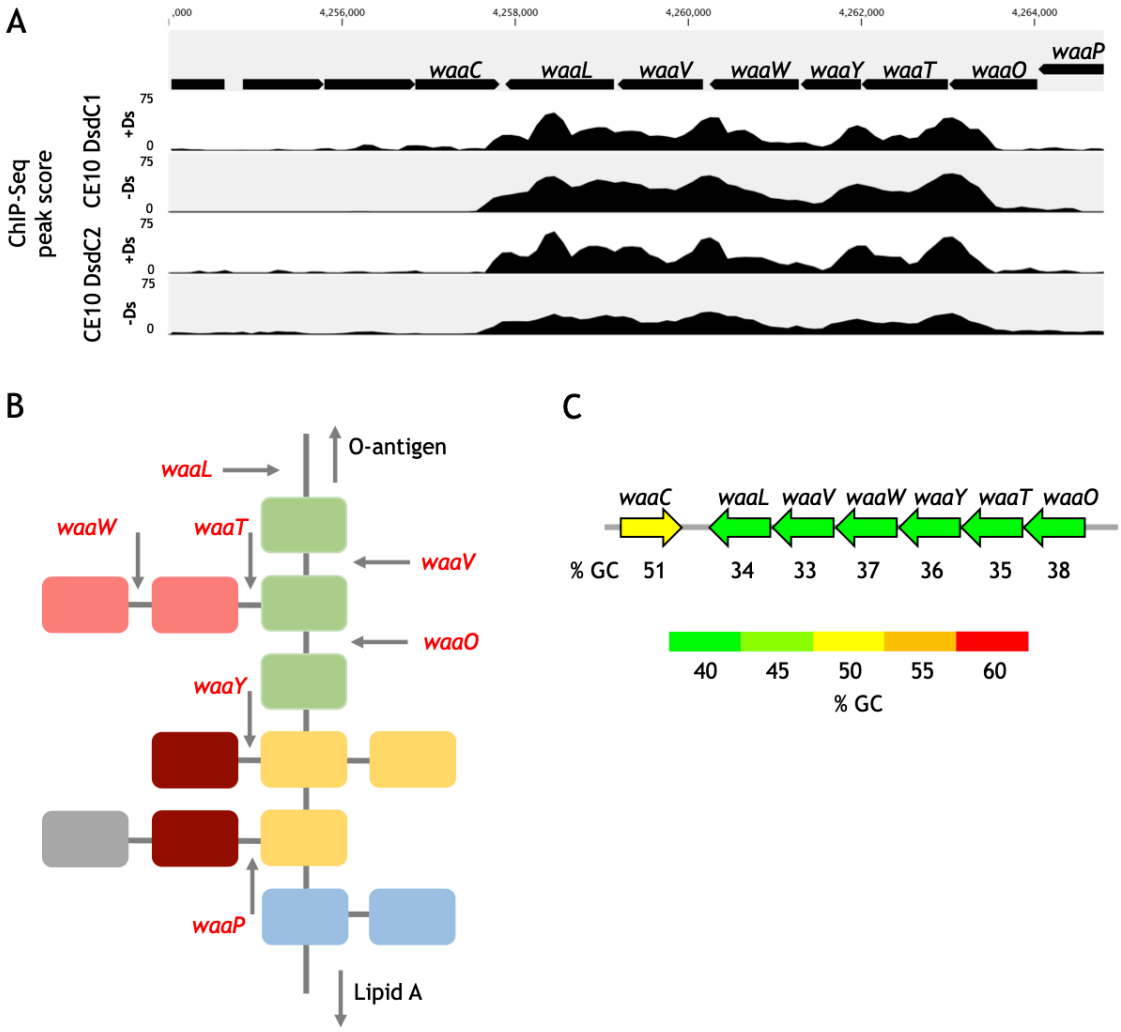


Figure 6-1 Expanded view of the ChIP-Seq peaks at the *waa* locus and representation of the *E. coli* LPS core OS structure. (A) An expanded view of the ChIP-Seq peaks from CE10^{DsdC1} and CE10^{DsdC2}, with and without D-serine, at the *waa* LPS core OS biosynthesis region. (B) Adapted from Yethon *et al.*, 2000. An overview of the sugar residues in the core OS and some of the core enzymes that modulate it. The blue rectangles indicate ketodeoxyoctonate; yellow rectangles indicate L-glycero-D-manno-heptose sugars; red rectangles indicate phosphates; grey rectangle indicates 2-aminoethyl phosphate; green rectangles indicate D-glucose and salmon rectangles indicate D-galactose sugar residues. (C) GC content of the *waa* locus genes in CFT073. BLAST analysis of the CFT073

waa locus to CE10 revealed 99% homology over 100% of the sequence. Adapted from xBase (Chaudhuri & Pallen, 2006).

Silver staining assays can be used to define LPS profiles, revealing any differences in core OS or O-antigen formation (Fomsgaard *et al.*, 1990; Tsai & Frasch, 1982). A silver staining assay was therefore performed to determine the LPS profiles of CE10 and the ΔΔ*dsdC1/2* mutant, in the presence of D-serine (Fig. 6-2).

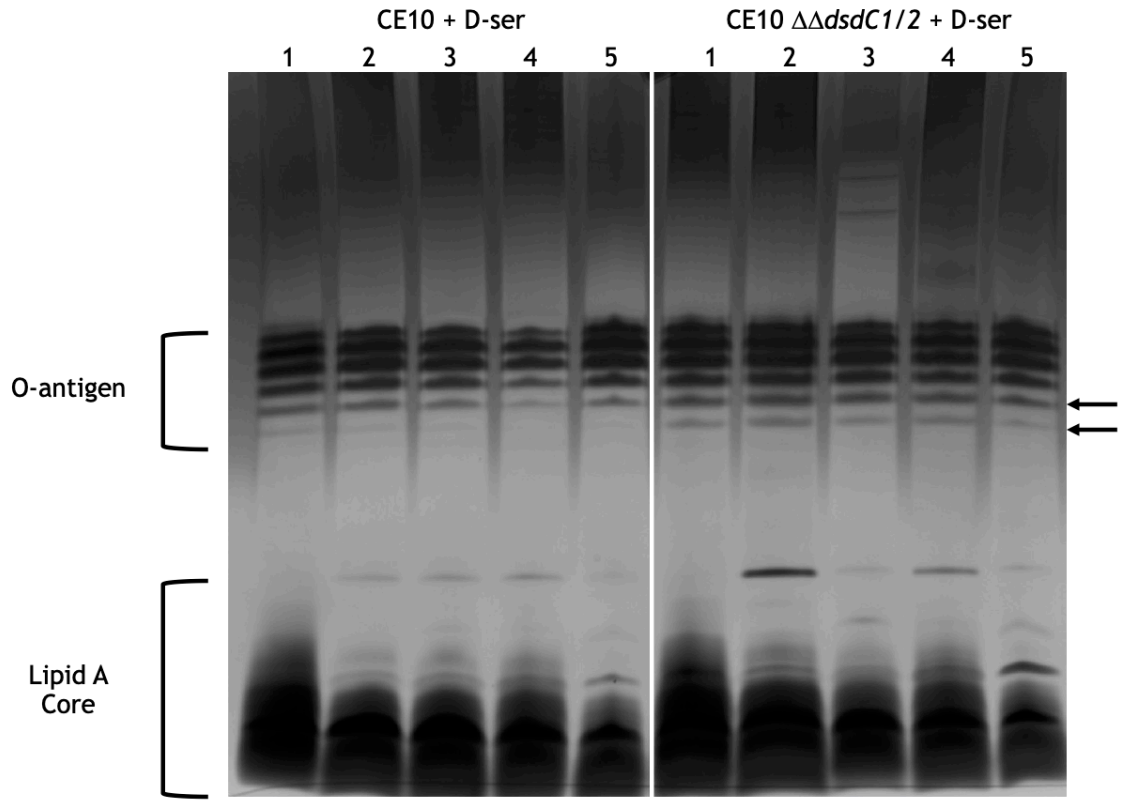


Figure 6-2 DsdC modifies the LPS of CE10. Silver-stained LPS profiles from five biological replicates of CE10 and ΔΔ*dsdC1/2*, in the presence of D-serine. Black arrows indicate differences in intensity in O-antigen repeat units.

Figure 6-2 showed the difference in LPS profiles between CE10 and ΔΔ*dsdC1/2*, in the presence of D-serine. The intense bands at the bottom of the gel represented the core OS and lipid A, and was present in all of the samples. The bands above the core represented the number of repeating O-antigen units in the samples. The ΔΔ*dsdC1/2* LPS profiles appeared to have two bands that were thicker than in the WT LPS profiles, indicated on Figure 6-2 by two black arrows. This difference in banding intensity suggests that DsdC may be playing a role in

modifying the LPS of CE10, potentially altering the number of O-antigen repeats that are present on the cell surface.

6.3 Further potential roles of DsdC in CE10 K1 capsule regulation

Further to regulating *neuO*, the capsule *O*-acetyltransferase, DsdC was observed binding to other genes involved in the synthesis of the K1 capsular polysaccharide in CE10 (Fig. 6-3A). Figure 6-3(A) highlighted the two types of binding patterns that were observed in the ChIP-Seq data. In the intergenic region before *kpsF* and *kpsM*, there was a single canonical TF binding peak. However, within Region 2 of the *kps* island, DsdC bound in a broad manner, encompassing many of the genes. Nucleotide analysis of the *kps* island in UTI89, another K1 *E. coli* strain, revealed that Region 2 was heavily A:T-rich, with the GC content comprising between 26-32% (Fig. 6-3B). A BLAST analysis of the entire *kps* pathogenicity island between UTI89 and CE10 revealed 99% homology over 94% of the sequence, indicating that Region 2 in CE10 is also A:T-rich.

As the ChIP-Seq data had revealed DsdC binding in the two major promoter elements of the K1 capsule pathogenicity island, RT-qPCR was performed on *kpsF* and *kpsM* to measure the changes in gene expression between CE10 WT and Δ*dsdC1/2*, in the presence and absence of D-serine (Fig.6-3C). Intriguingly, in the Δ*dsdC1/2* mutant, expression of *kpsF* was 1.7-fold higher (*p*-value 0.002) in relation to the CE10 WT (Fig. 6-3C). Upon exposure to D-serine, this increased to 2.5-fold higher (*p*-value 0.001) (Fig. 6-3C). This data implies that DsdC directly represses *kpsF*, the first gene in the Region 1 *kps* pathway. Furthermore, *kpsM* was expressed 4.6-fold higher in the Δ*dsdC1/2* mutant, in the presence of D-serine, in relation to the CE10 WT (Fig. 6-3C). However, in the absence of D-serine, there was no difference in expression of *kpsM* between CE10 WT and Δ*dsdC1/2*, suggesting accumulation of D-serine may have led to increase in *kpsM* gene expression described above.

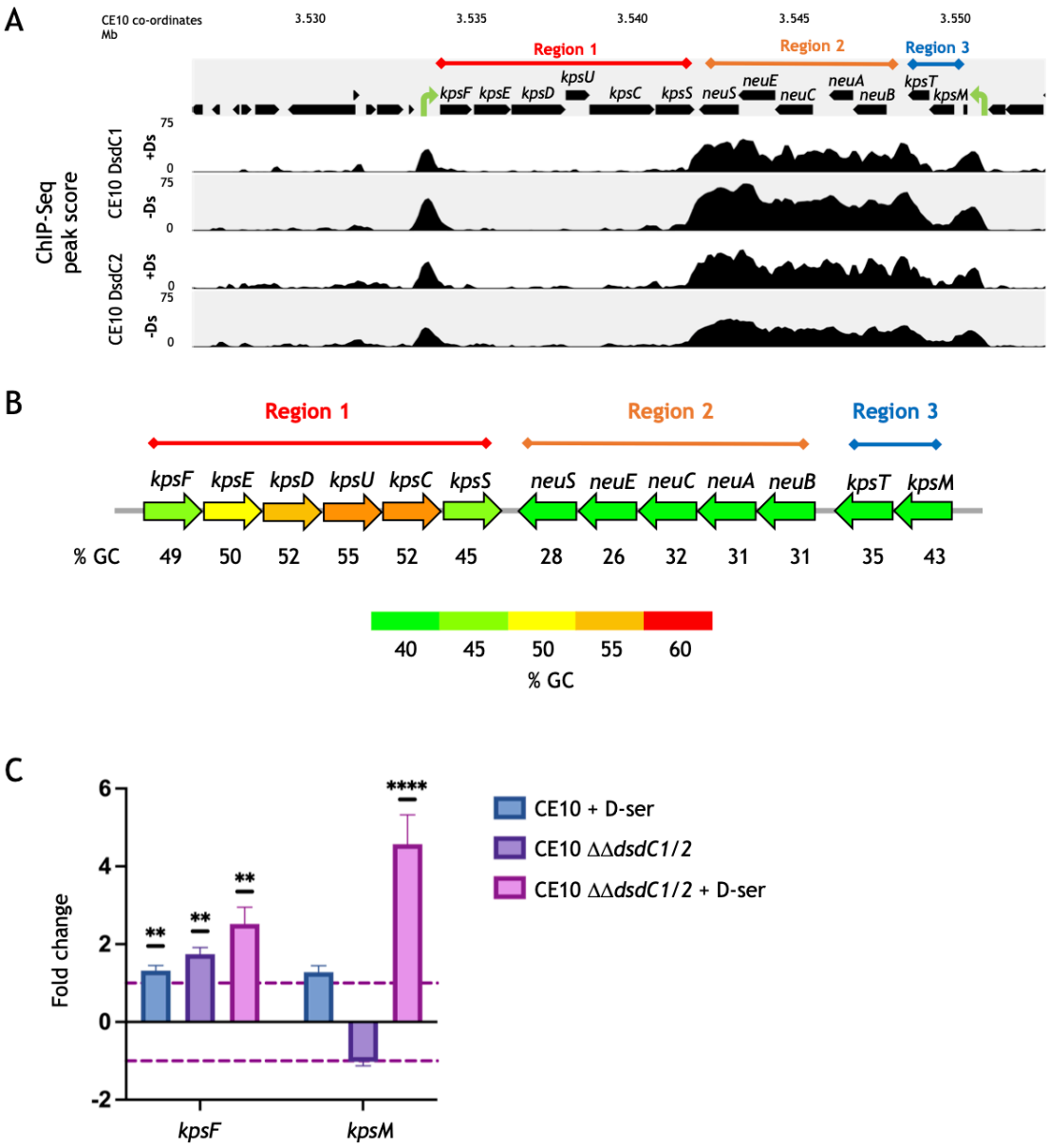


Figure 6-3 Expanded view of the ChIP-Seq peaks at the *kps* K1 capsular biosynthesis region and relative expression of *kpsF* and *kpsM*. (A) An expanded view of the ChIP-Seq peaks from CE10^{DsdC1} and CE10^{DsdC2}, with and without D-serine, at the *kps* K1 capsular biosynthesis region. The three genomic regions of the K1 capsule are annotated: Region 1 in red and inclusive of genes from *kpsF* to *kpsS*; Region 2 in orange and inclusive of genes from *neuS* to *neuB*; and Region 3 in blue, inclusive of the genes *kpsT* and *kpsM*. The two major promoter elements are represented by the green curved arrows, which indicate the direction of transcription, adapted from Jia *et al.*, 2017. (B) GC content of the *kps* locus genes in UTI89. BLAST analysis of the UTI89 *kps* locus to CE10 revealed 99% homology over 94% of the sequence. Adapted from xBase

(Chaudhuri & Pallen, 2006). (C) RT-qPCR analysis of the *kpsF* and *kpsM* K1 capsular genes. Fold changes were measured relative to the CE10 WT, with the purple dashed line indicating baseline expression of *kpsF* and *kpsM* genes in CE10 WT. * indicates a *p*-≤0.05; ** indicates a *p*-≤0.01; **** indicates a *p*-≤0.0001.

6.3.1 Adhesion and invasion of cultured brain microvascular endothelial cells does not appear to be mediated by DsdC

Although the full role of the NMEC K1 capsule has not been completely elucidated yet, it is thought to be involved in invasion and adhesion of the BBB in humans (Xie *et al.*, 2004). The adhesion and invasion capacities of CE10 and isogenic mutants were therefore measured. An *in vitro* brain microvascular endothelial cell line, hCMEC/D3, was used as a BBB model to measure invasion and adhesion levels of CE10 and isogenic mutants. Adhesion and invasion assays were carried out by incubating hCMEC/D3 cells with bacteria at an MOI of 100 for 2 hours. An MOI of 100 was used based upon prior NMEC *in vitro* research (Kim, 2000).

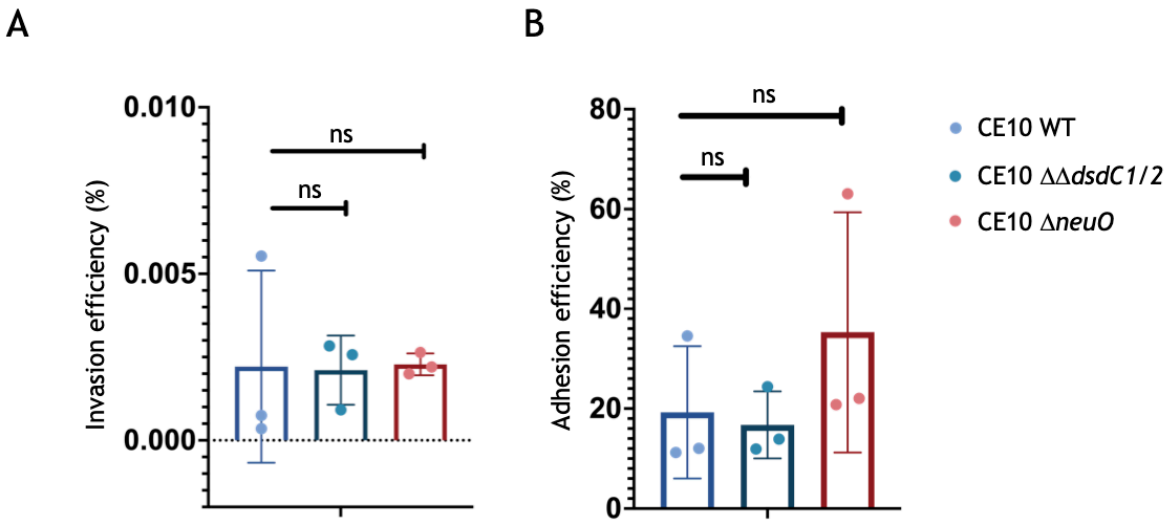


Figure 6-4 DsdC does not affect invasion or adhesion efficiency of CE10 in an *in vitro* BBB model cell line. Using the hCMEC/D3 brain microvascular endothelial cell line, the percentage of (A) invasion efficiency and (B) adhesion efficiency was measured in CE10, Δ*dsdC1/2* and Δ*neuO* against the total bacterial count (adherent, invasive and unattached). The hCMEC/D3 cells were incubated with bacteria at an MOI of 100, for 2 hours. Cells for the adhesion assay were lysed and bacterial counts performed. Cells for the invasion assay

were treated with gentamicin and grown for a further hour. Cells were lysed and bacterial counts performed. An unpaired Student *t*-test measurement was performed between the WT and isogenic mutants for each assay; *p*-≥0.05 were non-significant (ns). Data presented is from three independent biological repeats.

Of the total bacterial count, CE10 ΔΔ*dsdC1/2* displayed an invasion efficiency of ~0.002% to the hCMEC/D3 cell line. This was not significantly different to the percentage of invasion displayed by the WT, nor Δ*neuO* (Fig 6-4A). Furthermore, there was no apparent difference in the percentage of adhesion efficiency for ΔΔ*dsdC1/2* compared to the WT (Fig. 6-4B). Of the total bacterial count, ~20% of WT and ΔΔ*dsdC1/2* adhered to the hCMEC/D3 cell line, and ~36% of Δ*neuO*. These results indicate that any alterations to the polysialic acid K1 capsule, mediated by DsdC, did not affect the adhesion and invasion capabilities of CE10 in an *in vitro* cell line model.

6.3.2 DsdC does not appear to affect serum resistance

The K1 polysialic acid capsule is known to be involved in protection against host complement-mediated killing, providing a steric barrier to the cell (Miajlovic & Smith, 2014). To establish if modification of the K1 capsule by DsdC was involved in host complement-mediated killing, a serum resistance assay was performed on CE10 and Δ*dsdC* mutants, with K12 used as a serum sensitive positive control. Bacteria were incubated for 90 minutes with either PBS or human serum. Bacterial colony counts were then performed (Fig. 6-5).

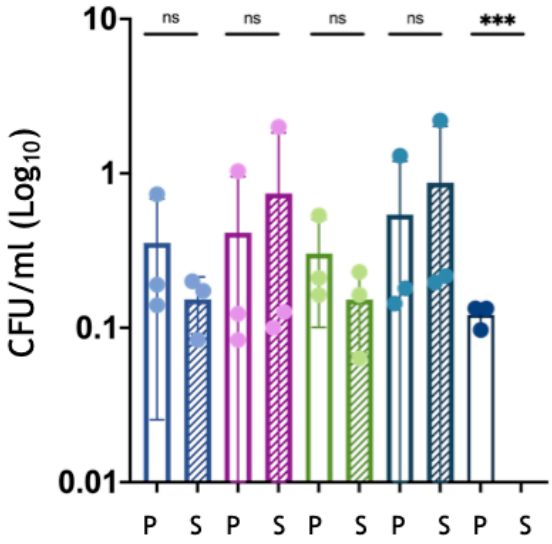


Figure 6-5 DsdC does not appear to affect serum resistance in CE10. Serum resistance was measured in K12, CE10, and CE10 isogenic mutants. Bacteria were incubated with PBS (P) or human serum (S) for 90 minutes, at a 1:1 concentration, and bacterial counts were performed (CFU/ml). (A) An unpaired Student *t*-test was performed on the K12, CE10 and CE10 isogenic mutant bacterial counts (CFU/ml (Log_{10})) between PBS incubation and serum incubation.

No significant difference was observed in the bacterial colony counts between CE10 WT incubated in PBS, compared to CE10 WT incubated in human serum (Fig. 6-5). This correlates with previous data that NMEC protects against host complement-mediated killing. Furthermore, no differences were observed in the bacterial colony counts in the isogenic mutants that were incubated in PBS compared to those incubated in human sera (Fig. 6-5). This implies that any modification of the capsule via regulation by DsdC, does not play a role in serum resistance.

6.4 The role of DsdC in regulating expression of genes involved in colibactin synthesis

Colibactin is a bacterial genotoxin that is produced by some strains of *E. coli* (Faïs *et al.*, 2018). Recent work in our lab has shown that D-serine can reduce the genotoxic effects caused by colibactin, however the mechanism for this remains unknown (Hallam *et al.*, unpublished). Intriguingly, the CFT073 ChIP-Seq data revealed binding of DsdC, upstream of the *clbR* gene (Fig. 6-6).

Furthermore, the RT-qPCR transcriptional data showed a 4.6-fold increase in expression of *clbR* in the Δ*dsdC* mutant plus D-serine background (*p*-value 0.0003) (Fig. 4-2A); indicating that DsdC, upon exposure to D-serine, represses c*lbR* and thus colibactin synthesis, thereby reducing the damaging genotoxic effects inflicted upon the cell. To understand the role of DsdC in colibactin synthesis, an *in vitro* genotoxin infection model was used.

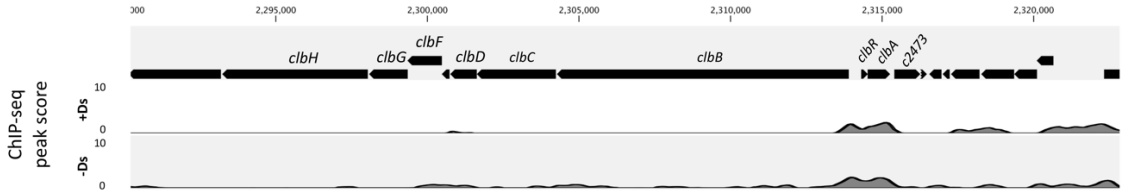


Figure 6-6 Expanded view of the *clbB-clbR* ChIP-Seq peak. An expanded view of the ChIP-Seq peak from CFT073^{DsdC}, in the presence and absence of D-serine, at the *clbB-clbR* intergenic region.

6.4.1 Experimental design and transcriptional analysis of colibactin genes

Colibactin synthesis can be measured *in vitro* using the eukaryotic model cell line, HeLa (Nougayrède *et al.*, 2006). Infection of HeLa cells by colibactin-producing *E. coli* results in DNA DSB, which can be measured indirectly through the phosphorylation of histone H2AX, and by the induction of megalocytosis, characterised by enlargement of the HeLa cell (Nougayrède *et al.*, 2006). As UPEC already encodes several different protein toxins (Welch, 2016), the model commensal strain for colibactin synthesis, Nissle 1917, was used to assess if DsdC was affecting colibactin synthesis. A nucleotide BLAST comparison revealed that CFT073 *dsdC* and Nissle 1917 *dsdC* were highly conserved with 100% conservation across 91% of the sequence. A Δ*dsdC* mutant was therefore generated in Nissle 1917. Growth profiles of Nissle 1917 and isogenic Δ*dsdC* mutant were performed to qualitatively assess growth in both M9 minimal media and MEM-HEPES, the media in which the *in vitro* infection assay would be performed. Furthermore, to ensure the Nissle Δ*dsdC* mutation had been successful, Nissle Δ*dsdC* was grown on 10 mM D-serine sole carbon source plates, with the expectation that it would not be able to metabolise D-serine and therefore no growth would be observed.

Moreover, *dsdC* was transformed into pACYC184 and complemented into Nissle Δ*dsdC*, with the expectation this would restore any growth defects observed in the Δ*dsdC* mutant.

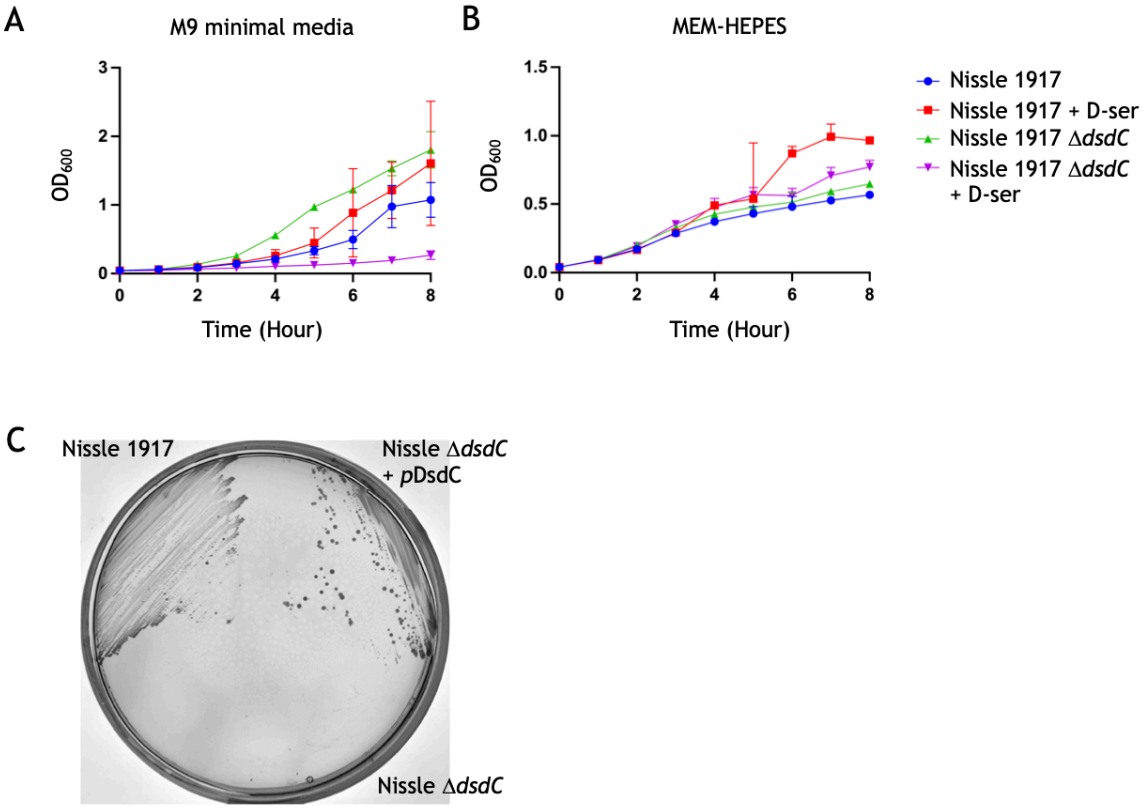


Figure 6-7 Growth profiles of Nissle WT and Δ*dsdC*. Growth curves of Nissle WT and Δ*dsdC* in (A) M9 minimal media and (B) MEM-HEPES media, supplemented with 1 mM of D-serine, measured as OD_{600} over time. Growth curves shown represent mean values of triplicate experiments with error bars indicating SEM. (C) Growth profile of Nissle WT, Δ*dsdC*, and Δ*dsdC* + pDsdC on sole carbon source plates supplemented with 10 mM D-serine.

In M9 minimal media, when D-serine was present, the Δ*dsdC* mutant was unable to grow (Fig. 6-7A/C). From this data, it was implied that without the DsdC regulator, D-serine metabolism could not occur in Nissle 1917. Furthermore, when pDsdC was complemented into the Δ*dsdC* mutant, the growth phenotype was restored (Fig. 6-7C). Intriguingly, in M9 minimal media without D-serine, upon deletion of the DsdC regulator, there was a qualitative difference in growth between Nissle WT and the Δ*dsdC* mutant, with the Δ*dsdC* mutant appearing to grow faster than the WT (Fig. 6-7A). This phenotype was not observed in MEM-HEPES, where Nissle WT and the Δ*dsdC* mutant grew

comparably (Fig. 6-7B). Upon exposure to D-serine, there was a slight difference in growth, with Nissle WT proliferating faster from hour 5 (Fig. 6-7B).

In Chapter 4, the expression of genes involved in colibactin synthesis had been measured using RT-qPCR in CFT073 (Fig. 4-2). This data revealed that *clbR* expression was 2-fold higher in the Δ*dsdC* mutant (*p*-value 0.0003), and 4.6-fold higher in the Δ*dsdC* mutant (*p*-value 0.0003) upon exposure to D-serine, in comparison to the CFT073 WT. As the *in vitro* infection would be performed with the Nissle 1917 strain, RT-qPCR was performed to measure the fold changes in gene expression of three genes involved in colibactin synthesis: *clbR*, the transcriptional regulator of colibactin synthesis; *clbA*, a phosphopantethinyl transferase; and *clbB*, a non-ribosomal peptide synthetase. The RT-qPCR was performed in M9 minimal media and MEM-HEPES on the Nissle WT and isogenic Δ*dsdC* mutant, in the presence and absence of D-serine.

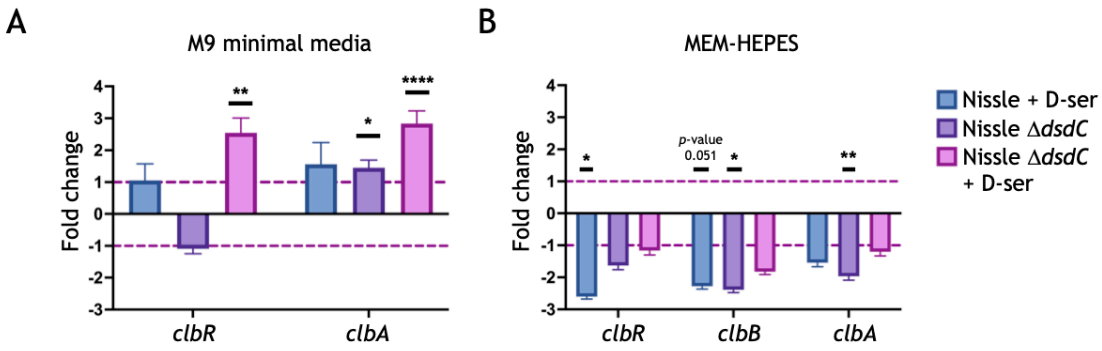


Figure 6-8 Relative expression of genes involved in colibactin synthesis. RT-qPCR analysis of genes involved in colibactin synthesis in (A) M9 minimal media and (B) MEM-HEPES for Nissle WT and Δ*dsdC*, in the presence and absence of D-serine. Fold changes were measured relative to the Nissle WT, with the purple dashed line indicating baseline expression of colibactin synthesis genes in Nissle WT. * indicates a *p*-≤0.05; ** indicates a *p*-≤0.01; **** indicates a *p*-≤0.0001.

Upon exposure to D-serine, expression of *clbR* in the Δ*dsdC* mutant in M9 minimal media, was 2.5-fold higher than in Nissle WT (*p*-value 0.005) (Fig. 6-8A). This correlates with the transcriptional data obtained earlier in CFT073 where the fold change was 4.6-fold higher in Δ*dsdC* compared to WT (*p*-value 0.0003) (Fig.4-2). However, in M9 minimal media without D-serine, changes in expression

of *clbR* in the Δ*dsdC* mutant were not significant in relation to the Nissle WT (Fig.6-8A). This was in contrast to the CFT073 transcriptomic data, in which expression of *clbR* in the Δ*dsdC* mutant was 2-fold higher compared to the CFT073 WT (*p*-value 0.0003) (Fig.4-2). Expression of *clbA*, the phosphopantethinyl transferase also required for production of colibactin, was 2.83-fold higher in the Δ*dsdC* mutant with D-serine in comparison to the Nissle WT (*p*-value 0.00002), which further correlates with the CFT073 data (Fig.6-8A/Fig.4-2). Expression of *clbA* was also 1.4-fold higher in the Δ*dsdC* mutant in the absence of D-serine (*p*-value 0.04). This indicates that upon exposure to D-serine in M9 minimal media, DsdC represses the expression of colibactin synthesis genes *clbR* and *clbA*. However, as *clbR* expression did not change in the Δ*dsdC* mutant and the change in *clbA* was only 1.4-fold increased, the changes in expression described above could be due to D-serine accumulation, rather than transcriptional effects mediated by DsdC.

The expression of colibactin synthesis genes were also measured in MEM-HEPES, as the *in vitro* infection cell assay would be performed in this media. This data revealed that there was no significant difference in expression of *clbR* between the Δ*dsdC* mutant and Nissle WT in MEM-HEPES, in the presence or absence of D-serine (Fig. 6-8B). This data indicated that DsdC is not regulating *clbR*. However, expression of *clbA* (-1.9-fold; *p*-value 0.007) and *clbB* (-2.4-fold; *p*-value 0.01) were both reduced in the Δ*dsdC* mutant in relation to the Nissle WT, indicating that DsdC is indirectly involved in the regulation of two colibactin synthesis genes. Expression of *clbB* was also -2.3-fold lower in the Nissle WT upon exposure to D-serine, in relation to the Nissle WT without D-serine (*p*-value 0.051), correlating to previous data obtained in our lab.

This data demonstrated that in different media, the expression of *clbA* and *clbR* changes in the Δ*dsdC* mutant compared to the Nissle WT. In MEM-HEPES, *clbR* is no longer differentially expressed by DsdC, indicating that DsdC does not regulate colibactin synthesis through the colibactin synthesis regulator. Thus, any differences therefore observed in the *in vitro* infection assay, would not be due to this regulatory mechanism. However, *clbA* and *clbB*, which are both also essential in colibactin synthesis, were repressed in the Δ*dsdC* mutant, suggesting that DsdC is involved in activation of colibactin synthesis in Nissle WT, in MEM-

HEPES. As the *in vitro* infection assay will be performed in MEM-HEPES, the transcriptional data indicates that the Δ*dsdC* mutation may potentially exacerbate any phenotypes observed.

6.4.2 DsdC does not appear to affect colibactin production in the *in vitro* infection assay

Previous work from our lab has revealed that upon exposure to D-serine, colibactin synthesis is reduced (Hallam *et al.*, unpublished). This was measured transcriptionally through the reduced expression of *clbB* in Nissle WT upon addition of D-serine. Furthermore, this was indirectly measured *in vitro* through Western blot analysis of the phosphorylated γH2A.X, a sensitive indicator of DNA DSB, and by the induction of megalocytosis ("giant cell phenotype") in the HeLa cells. To understand the potential role of DsdC in colibactin synthesis, an *in vitro* infection assay was performed on Nissle WT and the Δ*dsdC* mutant, in the presence and absence of D-serine. Briefly, the megalocytosis phenotype was observed by infecting HeLa cells with bacteria at an MOI of 400 for 4 hours. The HeLa cells were then treated with gentamicin and incubated for a further 48 hours, to elicit the genotoxic effect of colibactin. To measure, indirectly, the DNA DSB, the HeLa cells were infected with bacteria at an MOI of 400 for 4 hours. The HeLa cells were then treated with gentamicin and incubated for a further 4 hours, before the cells were removed for Western blot analysis. The *in vitro* infection experiments were completed in co-ordination with a fellow PhD student of the Roe group, Jennifer Hallam, who then performed the Western blot analysis.

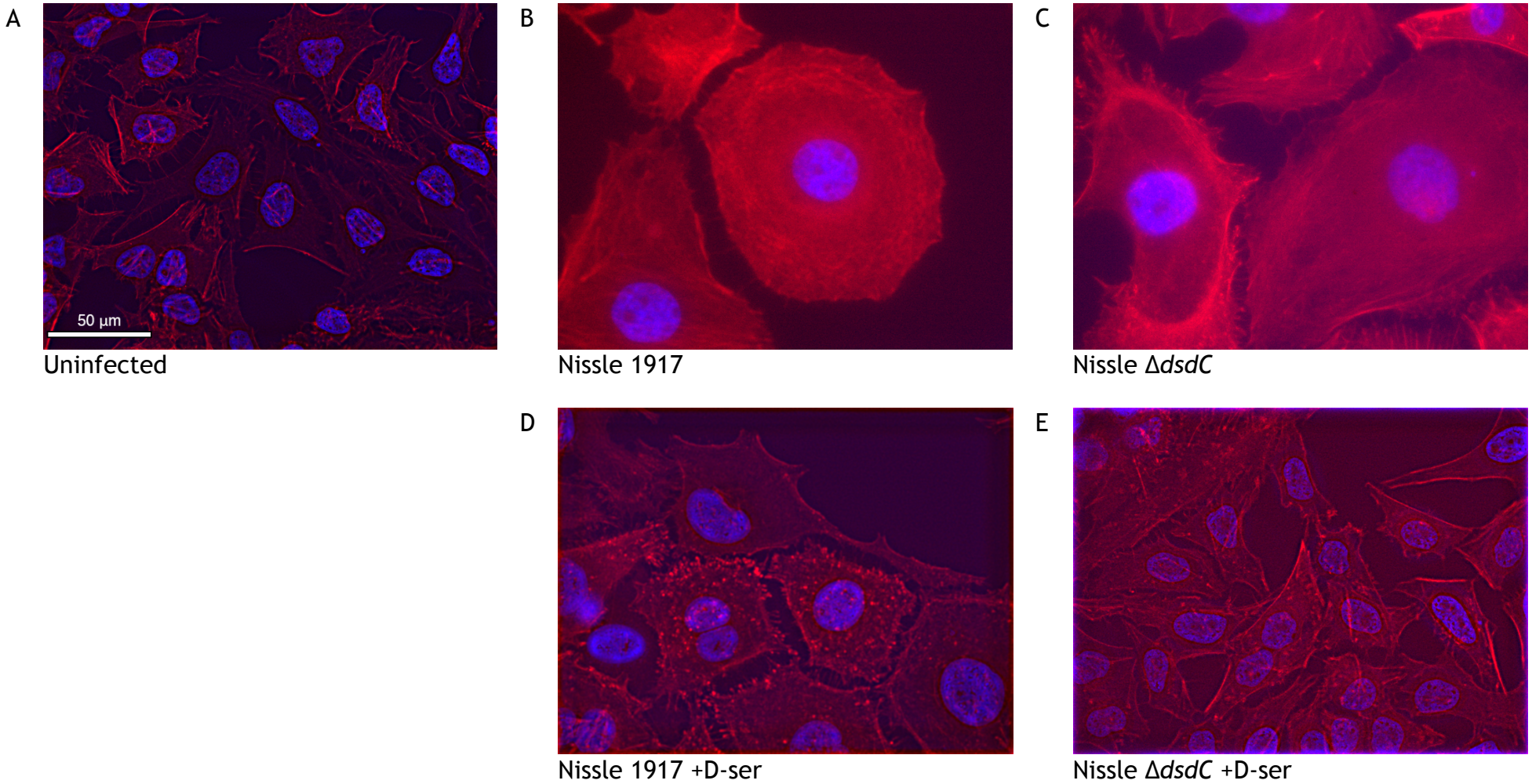


Figure 6-9 DsdC does not affect induction of megalocytosis in HeLa cells.

Figure 6-9 DsdC does not affect induction of megalocytosis in HeLa cells. Using the HeLa model cell line, colibactin-induced megalocytosis was observed microscopically. HeLa cells were incubated with bacteria at an MOI of 400. Images are representative immunofluorescence images of HeLa cells, 48 hours after infection with (B) Nissle 1917, (C) Nissle Δ*dsdC*, (D) Nissle 1917 + D-ser, (E) Nissle Δ*dsdC* + D-ser and (A) uninfected. Filamentous-actin was stained with Phalloidin AlexaFlour 555 (red), and host nuclei stained with DAPI (blue). Images were taken at 40X magnification.

Figure 6-9 shows the genotoxic effects of colibactin on the host cell. Infection of the HeLa cells by both Nissle and the Δ*dsdC* mutant caused the megalocytosis phenotype that is characteristic of colibactin production (Fig. 6-9B/C). No differences were observed between WT and Δ*dsdC*. Moreover, exposure to D-serine to both Nissle WT and Δ*dsdC* infections, produced a cell phenotype more characteristic of the uninfected cells (Fig. 6-9A/D/E). This correlated with previous data from our lab, showing that addition of D-serine to Nissle inhibits colibactin production. The Δ*dsdC* mutant did not appear to show a difference in colibactin production, measured qualitatively through the megalocytosis phenotype, in comparison to the Nissle WT, with or without D-serine.

As colibactin can induce DNA DSB in host cells, using the *in vitro* infection cell model, the concentration of phosphorylated histone γ-H2A.X was measured using Western blot analysis.

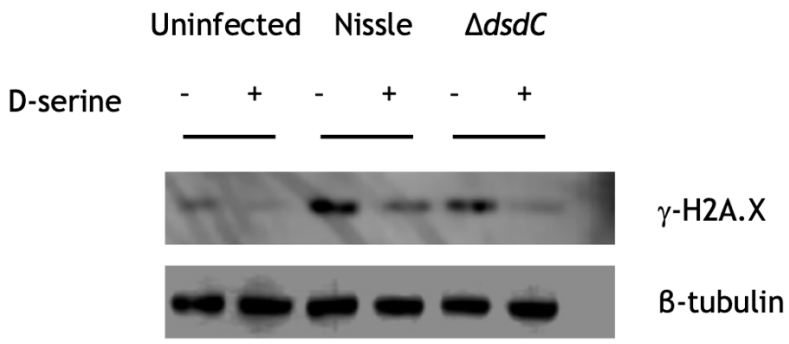


Figure 6-10 DsdC does not appear to affect induction of host DNA DSB. Western blot analysis of phosphorylated histone, γ-H2A.X, in both uninfected HeLa cells and HeLa cells infected with Nissle and Δ*dsdC*, in the presence and

absence of D-serine. β-tubulin was used as a loading control. This figure is representative of 3 biological repeats.

Phosphorylated histone γ-H2A.X is a sensitive marker of DNA DSB. Infection of both the Nissle WT and Δ*dsdC* mutant on the HeLa cells revealed an increase in phosphorylated γ-H2A.X in relation to the uninfected cells, indicating DNA DSB have occurred within the cell (Fig. 6-10). Over three biological experiments, Western blot analysis revealed no difference between Nissle WT and the Δ*dsdC* mutant in regard to γ-H2A.X expression (Fig. 6-10). Upon exposure to D-serine, there was a clear reduction in γ-H2A.X in both the Nissle WT and Δ*dsdC* mutant, indicating that D-serine was repressing the genotoxic effects of colibactin. In the presence of D-serine, there did not appear to be a difference between the Nissle WT and Δ*dsdC* mutant in regard to γ-H2A.X expression (Fig. 6-10). These data suggest therefore, that although there appeared to be direct binding and regulation of DsdC to *clbR* in the CFT073 background, this did not translate to a phenotype that showed a difference in the genotoxic effects of colibactin.

6.5 Discussion

6.5.1 DsdC modifies the O-antigen length of CE10 LPS

A key virulence factor of *E. coli* is the LPS, which constitutes a major component of the outer cell envelope, covering ~75% of the cell surface area (reviewed in Lerouge & Vanderleyden, 2002). It is the O-antigen region of the LPS molecule that is exposed to the host environment, thus variation in O-antigen composition is an important mechanism for the bacteria to evade the host immune system (reviewed in Liu *et al.*, 2020). Analysis of the ChIP-Seq binding data revealed occupancy of DsdC along the full *waa* operon, the major operon responsible for core LPS biosynthesis (Fig. 6-1A). Nucleotide analysis of this region revealed it was A:T-rich, with a GC content comprising between 33-38% of the sequence (Fig. 6-1C). The GC content amongst all strains of *E. coli* is relatively constant, ~50% (Muto & Osawa, 1987); with CE10-specific GC content at 50.6% (Liu *et al.*, 2015). Differences in GC content in the genome can be due to the acquisition of "foreign" DNA by horizontal gene transfer (Mann & Chen, 2010). NAPs, for example H-NS, can bind to A:T-rich regions of the genome as a mode of gene regulation, silencing "foreign" genes (Grainger *et al.*, 2006). As discussed in

section 3.6.5, DsdC has many similarities to NAPs. As the *waa* operon is A:T-rich, in comparison to the GC content of the CE10 genome, DsdC may be occupying the *waa* operon in a NAP-like manner, ensuring silencing of the operon. This correlates with the transcriptomic data which revealed expression of genes involved in LPS core OS biosynthesis (*waaV, waaY, waaL, waaW,* and *waaT*) were increased in the ΔΔ*dsdC1/2* mutant in relation to the WT, in the presence of D-serine, suggesting that DsdC directly represses the *waa* operon.

As discussed previously, the *waa* operon is the major operon responsible for core OS biosynthesis (Heinrichs *et al.*, 1998). Elongation of the LPS molecule is tightly regulated, and the assembly of sugars is mediated by a host of glycosyltransferases (Whitfield *et al.*, 1999). One of the most significant DsdC binding peaks in the *waa* operon was in the region of a glycosyltransferase-encoding gene, *waaV* (Table S3-6). WaaV has been shown to link the final sugar to the core region before the O-antigen is added by the O-antigen ligase, *waaL* (Yethon *et al.*, 2000). Mutations of *waaV* have been shown to eliminate the ligation of the O-antigen, revealing its importance in LPS formation (Heinrichs *et al.*, 1998). Using EMSAs, direct binding of DsdC to the *waaV* gene was demonstrated (Fig. 4-7). Furthermore, RT-qPCR showed a 1.36-fold increase in *waaV* expression in the ΔΔ*dsdC1/2* mutant compared to the WT (Fig. 4-1). This suggests that DsdC represses the *waaV* gene, through direct binding. Speculatively, the repression of *waaV* could cause a difference in the number of O-antigen repeat units. Silver staining of LPS from CE10 and the ΔΔ*dsdC1/2* mutant, in the presence of D-serine, revealed two bands of increased intensity in the O-antigen side chain region in the ΔΔ*dsdC1/2* mutant. This implies that in the WT, DsdC inhibits the formation of two O-antigen polysaccharides of varying lengths.

Alterations of the LPS and O-antigen can occur in response to different host environmental stimuli (Lerouge & Vanderleyden, 2002). Indeed, Van Den Akker, (1998) reported that temperature affected expression of the O-antigen in *Bordetella parapertussis*. Furthermore, pathogenic *Yersinia enterocolitica* were shown to modulate their LPS profile upon exposure to differences in pH and Ca^{2+} restriction (Bengoechea *et al.*, 2003). Intriguingly, the length of the O-antigen has been shown to affect complement resistance and evasion of macrophage

cells in *Salmonella* (Murray *et al.*, 2006). Hypothetically therefore, it could be speculated that DsdC, upon exposure to the host metabolite D-serine, directly represses *waaV* and the *waa* operon, causing alterations in the LPS O-antigen length as a defence mechanism of immune evasion. Although it has not been definitively proven if direct regulation of *waaV* by DsdC causes the difference in O-antigen length, it does appear using the silver staining data that DsdC, upon exposure to D-serine, modifies the composition of the LPS, a key virulence determinant in *E. coli*.

6.5.2 Further putative regulatory roles of DsdC in CE10 K1 capsule regulation

E. coli is one of the leading causes of blood-stream infections, with bacteraemia-causing strains often associated with serum resistance (Miajlovic & Smith, 2014). Human serum contains over thirty proteins involved in the complement-killing system, one of the primary defences in the host's immune response (Miajlovic & Smith, 2014). The K1 capsule is a known virulence factor that protects against complement-mediated killing (reviewed in Abreu & Barbosa, 2017). Analysis of the CE10 ChIP-Seq data revealed that DsdC bound in a broad manner in the A:T-rich Region 2 of the *kps* pathogenicity island (Fig. 6-3A/B), the island responsible for K1 capsule synthesis (Silver & Vimr, 1990). This suggests, as discussed above, that DsdC may potentially be acting in a similar manner as a NAP, and silencing "foreign" DNA. Furthermore, DsdC was shown to bind in the two major promoter regions of the *kps* island, upstream of *kpsF* and *kpsM* (Fig. 6-3A). RT-qPCR analysis revealed expression of the capsular gene, *kpsF*, was higher in the Δ*dsdC1/2* mutant in relation to the WT, implying that DsdC represses *kpsF*, the first gene in the Region 1 operon (Fig. 6-3C).

To determine if binding and differential expression of capsular genes by DsdC affected protection from complement-mediated killing, a serum resistance assay was performed on CE10 and the isogenic Δ*dsdC* mutants. As expected of the CE10 WT with a functioning capsule, there was no difference in bacterial growth between samples incubated in PBS and samples incubated in human serum, whereas the K12 serum sensitive strain was fully killed (Fig.6-5) (Leying *et al.*, 1990). Furthermore, the serum resistance assay revealed no difference in bacterial growth of the isogenic Δ*dsdC* mutants between the samples incubated

in PBS and the samples incubated in human serum, suggesting that the capsule is intact in the Δ*dsdC* isogenic mutants (Fig. 6-5).

As discussed above, DsdC was shown to repress the first gene in Region 1 of the *kps* island, *kpsF*. KpsF functions as an accessory translocation factor for capsular components, with Δ*kpsF* mutations reducing capsular expression in *E. coli* (Cieslewicz & Vimr, 1997). Serum resistance has been linked to the proportion of K capsule present, with strains producing more K-antigen having an enhanced resistance to serum (Glynn & Howard, 1970). This suggests that the Δ*dsdC* mutant should have enhanced survival in serum. However, analysis of serum survival in Δ*dsdC* mutants showed no significant difference. Conversely, Vermeulen *et al.*, (1988) suggested that there is a threshold level of K1 capsular polysaccharide that is needed to confer protection from complement-mediated killing. Hypothetically therefore, increased expression of *kpsF* in the ΔΔ*dsdC1/2* mutant may not have affected serum resistance, as the cell is already expressing the maximal amount of capsule needed to confer protection. Conversely, other capsular genes (*neuB, neuE, neuC, neuA,* and *neuS*) were only upregulated in the ΔΔ*dsdC1/2* mutant, in relation to the WT, in the presence of D-serine. Therefore, a difference in serum resistance may have been seen if the samples had been exposed to D-serine during the experiment. Indeed, exposure to environmental factors have been known to modify *E. coli* K1 resistance to serum, with exposure to 0.2 M NaCl eliciting a 3.3-fold decrease in survival (Badger & Kim, 1998). Although an interesting hypothesis, this experiment was not carried out in this project, but could be the subject of future work.

Furthermore, the K1 capsule is also one of many determinants that is required for successful invasion of the BBB (reviewed in Xie, Kim & Kim, 2004). Indeed, 84% of 77 isolates recovered from the CSF of neonates with meningitis had the K1 polysaccharide (Robbins *et al.*, 1974). An *in vitro* model cell line was therefore used to indirectly assess the role of DsdC on adherence and invasion of the BBB, through adhesion and invasion of brain endothelial cells. In the *in vitro* model, no difference in adhesion and invasion of CE10 and the ΔΔ*dsdC1/2* mutant was observed. Strikingly however, invasion to the hCMEC/D3 cells was low, with the bacteria showing an invasion efficiency of ~0.002% of the total bacterial cells. In comparison, an *in vitro* study that used HBMEC cells, *E. coli* K1

were shown to invade routinely at an efficiency of 0.1%-0.5% (Badger *et al.*, 2000). Indeed, a non-invasive K12 was used as a negative control and was shown to bind to the HBMEC cells at a similar percentage as the data presented here, ≤0.002% (Badger *et al.*, 2000). This insinuates that the CE10 WT is not efficiently invading the hCMEC/D3 cell line, thus no differences could be observed between WT and Δ*dsdC1/2*. It should be noted however, that Badger *et al.*, (2000) based their invasion efficiency percentages off the initial inoculum used, unlike the data presented here which was based off the total number of bacteria in the experiment (adherent, invasive and unattached bacteria).

The BBB is a dynamic interface between the brain and vasculature, composed of endothelial cells and connected by tight junctions (reviewed in Jamieson, Searson & Gerecht, 2017). Transwell models are often employed in *in vitro* studies, in order to mediate the dynamic nature of the BBB (reviewed in Stone, England & O'Sullivan, 2019). HBMEC confluent monolayers are grown on a porous membrane (transwell) that separates apical and basolateral chambers (Jamieson *et al.*, 2017). Addition of further cells involved in the maintenance of the BBB, including astrocytes and pericytes, can be added to the basolateral chamber, upregulating barrier function and allowing for a more dynamic BBB *in vitro* cell model (Jamieson *et al.*, 2017). For the *in vitro* invasion assays presented in this work, transwells were not used, and hCMEC/D3 cells were grown to a confluent monolayer in collagen coated wells. Potentially therefore, invasion efficiency was low because the *in vitro* cell model cell line was not optimised. Further studies could use transwell inserts to optimise CE10 WT binding before further studying the difference in invasion efficiency between CE10 and Δ*dsdC1/2*.

There was also no difference in adhesion and invasion between CE10 and Δ*neuO*. This corresponds with data obtained with NMEC strain RS218, which also showed no dependency on NeuO for adherence or invasion of HBMEC cells (Mordhorst *et al.*, 2009). Therefore, it appears from this data that neither DsdC nor NeuO plays a role in invasion of CE10 into hCMEC/D3 cells. However, invasion efficiencies were so low for CE10 WT in the *in vitro* BBB cell model, further studies would need to be performed in order to fully rule out the role of DsdC in K1 capsule invasion.

6.5.3 DsdC does not mediate the reduction in colibactin synthesis, induced upon exposure to D-serine, in Nissle 1917

Previous work in our lab had revealed that, upon exposure to D-serine, the genotoxic effects to eukaryotic cells mediated by the bacterial genotoxin, colibactin, was reduced (Hallam *et al.*, unpublished). Although this was an exciting discovery, the molecular mechanism underpinning the reduction in colibactin synthesis, mediated by D-serine, remained unknown. Analysis of the CFT073^{DsdC} ChIP-Seq data had revealed binding on the *pks* island, at the intergenic region between *clbR*, the colibactin transcriptional activator and *clbB* (Fig. 6-6). Transcriptional analysis revealed that expression of *clbR* was 2-fold higher in the CFT073 Δ*dsdC* mutant in relation to the WT and moreover, upon exposure to D-serine, this increased to 4.6-fold higher (Fig. 4-2). This indicated that in CFT073, DsdC represses *clbR* and thus colibactin expression. Furthermore, upon exposure to D-serine, this repression was exacerbated, and thus theoretically could explain how the induction of D-serine causes a repression of colibactin synthesis.

As UPEC encodes multiple toxins (Shah *et al.*, 2019; Welch *et al.*, 2002; Welch, 2016), the probiotic Nissle 1917 is used as the model UPEC strain in *in vitro* assays. Indeed, a recent analysis revealed few genetic differences between CFT073 and Nissle 1917, with only 308 CFT073 genes absent in Nissle 1917 (Vejborg *et al.*, 2010). A nucleotide BLAST analysis between CFT073 and Nissle 1917 revealed *dsdC* is highly conserved between the two strains. A Nissle Δ*dsdC* mutant was therefore made to assess the effects of DsdC on colibactin synthesis *in vitro*. Intriguingly it was qualitatively observed in a growth profile, in M9 minimal media, that the Nissle Δ*dsdC* mutant proliferated faster than the Nissle WT. This potentially implies that DsdC is repressing a gene involved in metabolism in Nissle. This phenotype was not present in the CFT073 Δ*dsdC* mutant and potentially shows strain-specific regulation of carbon sources between CFT073 and Nissle. Although this was an interesting observation, this was not investigated further due to time constraints.

Conversely, unlike the CFT073 transcriptomic data, in the Nissle Δ*dsdC* there was no difference in gene expression of *clbR* in relation to the Nissle WT,

suggesting that DsdC does not repress colibactin synthesis in Nissle. However, upon exposure to D-serine, *clbR* was repressed by DsdC in M9 minimal media. As LTTR need a co-inducer in order to differentially regulate genes, perhaps this was why a difference was seen in *clbR* expression in the presence of D-serine (reviewed in Maddocks & Oyston, 2008). Conversely, perhaps the difference in *clbR* expression in the Δ*dsdC* mutant was due to the inability to break down D-serine. Therefore, it is the negative effects associated with D-serine accumulation that is causing *clbR* to be differentially expressed. Strikingly though in MEM-HEPES, *clbR* expression was no longer significantly differentiated in the mutant, with or without D-serine. Different medias have been known to affect gene expression in bacteria (Ray *et al*., 2009). Indeed in *S. aureus*, it was shown that transcription of *hla* (α-haemolysin) was reduced in CYGP media compared to BHI or LB media (Ray *et al*., 2009). Furthermore, even slight changes in the carbon content of medias can vastly influence bacterial gene expression. In *Streptococcus mutans*, genes related to biofilm formation (*atlA*, *sacB*, and *wapA*) were significantly upregulated in 1% xylitol-supplemented media in comparison to media containing 0.58% glucose (Decker *et al*., 2014). These studies highlight the complexity of researching gene regulation and how the different components of media can vastly affect gene expression. Overall, from this data however, it appeared that DsdC does not regulate colibactin synthesis through the transcriptional activator, *clbR,* in MEM-HEPES.

Intriguingly in MEM-HEPES, RT-qPCR analysis revealed *clbA* and *clbB* were downregulated in the Nissle Δ*dsdC* mutant. This suggested that DsdC is indirectly activating the expression of colibactin. This did not however translate into a phenotype in the *in vitro* infection assay. Western blot analysis of a marker of DNA DSB did not show a difference between Nissle WT and Δ*dsdC* mutant, in the presence or absence of D-serine, over three biological replicates. Furthermore, no differences were observed in the megalocytosis phenotype between Nissle WT and Δ*dsdC* mutant, suggesting that DsdC does not play a significant role in the regulation of colibactin in Nissle.

6.5.4 Conclusion

Exposure to environmental signals can cause rapid changes in bacterial gene expression. The work presented in this chapter revealed that upon exposure to D-serine, DsdC mediated changes in the O-antigen length of the LPS, potentially as a mechanism for immune evasion. This adds to the evidence presented in this thesis that DsdC plays a larger role in gene regulation in CE10 and does not solely regulate the *dsdCXA* locus. Further, although no differences were observed between the CE10 WT and ΔΔ*dsdC1/2* mutant in serum resistance and invasion to the *in vitro* BBB model cell line, the possibility that DsdC modulates properties of the K1 capsule cannot be ruled out, and further experiments are required for conclusiveness.

Chapter 7 Final conclusions

E. coli is estimated to kill more than 2 million humans per year, posing a serious clinical threat to the world (reviewed in Tenaillon *et al.*, 2010). With extraintestinal infections on the rise (Vihta *et al.*, 2018), as well as the impending anti-microbial resistance (AMR) crisis, it is essential that we understand the fundamental processes that govern bacterial survival. CFT073 and CE10 are both prototypical UPEC and NMEC strains, that elicit disease in the urinary tract, and the brain respectively. Both UPEC and NMEC are able to use D-serine as a carbon source, and thus DsdC plays an integral role in the success of these two ExPEC strains. Tight regulatory control of genes in response to fluctuating metabolic and environmental signals is central to bacterial success. Indeed, *E. coli* has been reported to dedicate up to 6% of its genome on TFs, reportedly encoding over 300 DNA-binding regulatory proteins (Pérez-Rueda & Collado-Vides, 2000). Although, Anfora *et al.*, 2007 speculated a larger role in gene regulation for DsdC, to our knowledge this has yet to be ascertained. Therefore, there were three main questions that this work aimed to elucidate. Firstly, what was the direct regulon of DsdC in two *E. coli* pathotypes? Secondly, was there any significance in CE10 carrying two copies of the *dsdCXA* locus? And lastly, did DsdC-dependent regulation affect the physiology of either CFT073 or CE10.

Using a combination approach of ChIP-Seq and RNA-Seq, the direct regulon of DsdC was elucidated. It was revealed that DsdC bound to 217 regions of the CE10 genome and 129 regions of the CFT073 genome, indicating that DsdC plays a global role in *E. coli* regulation. Upon cross-examination with the RNA-Seq and RT-qPCR data, DsdC was found to directly regulate the colibactin master regulator *clbR*, in CFT073, and LPS and K1 capsular biosynthesis genes in CE10. Intriguingly, this revealed that DsdC has been adapted for strain-specific regulation within two distinct ExPEC pathotypes. The genetic diversity between *E. coli* strains is often ascribed to the relative plasticity of the genome, with some studies suggesting that up to 75% of genes were acquired through HGT (Touchon *et al.*, 2009; reviewed in Juhas, 2015). However, the *E. coli* genome is finite, and multiple selective pressures often dictate the acquisition or loss of genes, leading to vast phenotypic differences between strains of the same species (reviewed in Perez & Groisman, 2009b). Indeed, Welch *et al.*, 2002 described the relative mosaic fluidity of the *E. coli* genome; revealing that

between K12, UPEC, and EHEC, only 39.2% of combined proteins were shared between the three *E. coli* strains. Many characteristic virulence factors are encoded on PAI and thus were likely acquired through HGT. However, newly acquired genes are only beneficial if they are expressed at the optimum time, thus requiring a need to be integrated into the ancestral regulatory network of the cell (Perez & Groisman, 2009b). Ancestral TFs have therefore been discovered to be repurposed and tailored to regulate newly acquired genes (Perez & Groisman, 2009a; Connolly *et al.*, 2019). The results presented here indicate that is the case for DsdC, regulating distinct virulence factors in two strains of *E. coli*. Therefore, these findings further support that the common assumption that TFs regulate the same network within strains of the same species is not always correct, and TF studies using "model" laboratory strains should be interpretated with care.

The role of gene duplication in driving regulatory network evolution cannot be understated (Teichmann & Babu, 2004). Duplications are often pathotype specific, and are typically associated with the adaption of cells to a fluctuating environment (Bernabeu *et al.*, 2019). In 2006, Moritz and Welch, revealed that 30 out of 41 K1 strains were reported to carry two copies of the *dsdCXA* operon, with both copies able to catabolise D-serine, indicating that they are functionally redundant. This suggests an importance of duplication of the operon to K1 strains. Amino acid BLAST analysis revealed that in CE10, DsdC1 and DsdC2 were 98% homologous over 98% of the sequence. When genes evolve by duplication, regulatory networks are either conserved, lost or new interactions are gained (Teichmann & Babu, 2004). From the CE10 ChIP-Seq binding data, it appeared that the regulatory network between DsdC1 and DsdC2 had been conserved, as there were no differences observed between the two TFs. Furthermore, there appeared to be no demonstratable observable differences in D-serine metabolism between either copy, nor any difference in expression levels. However, genes with identical functions are often not stably maintained within the genome, unless the additional expression of gene product is advantageous (Zhang, 2003); leading to the conclusion that high levels of DsdC must be physiologically important for CE10. Indeed, both DsdC1 and DsdC2 were more abundant when compared to CFT073 DsdC, when native levels of DsdC were compared by Western blot analysis in Chapter 3.

While functionally redundant in their shared capacity to activate D-serine catabolism and appearing to share similar genomic binding sites, a distinction in regulatory potential between DsdC1 and DsdC2 was revealed during an analysis of *neuO*, a capsular *O*-acetyltransferase encoding gene that appeared to be bound by both DsdC1 and DsdC2. The capsule is a key virulence determinant for *E. coli* and is the first line of defence against the host immune system. Although there are over 70 capsular antigens, the K1 polysialic acid capsule is often over-represented in NMEC infections (Robbins *et al.*, 1974; Jann & Jann, 1992), indicating its importance in NMEC survival. Indeed, other pathogens that are associated with neonatal meningitis, such as *K. pneumoniae* and *N. meningitidis*, commonly possess a similar polysialic acid capsule (Opal, 2014; Stephens *et al.*, 1993). *O*-acetylation of the polysialic acids that comprise the capsule occurs through the phase variable gene, *neuO* (Deszo *et al.*, 2005). Although *O*-acetylation of the *N. meningitidis* capsule has been implicated in virulence of the strain, the role of *O*-acetylation of the *E. coli* K1 capsule has been more obscure, with conflicting reports on its role in virulence (Frasa *et al.*, 1993; Colino & Outschoorn, 1999; Fusco *et al.*, 2007).

As *neuO* was originally described as being encoded on the prophage CUS-3, a logical hypothesis would suggest that *O*-acetylation of the capsule modifies the bacteriophage receptor, thereby preventing further phage entry and stopping bacteriophage "super-infections" (Deszo *et al.*, 2005; Labrie, Samson & Moineau, 2010). In this work it was speculated that CUS-3 has been lost in CE10, with sequence homology between the incomplete prophage in CE10 and CUS-3 revealing potential degradation. Conversely, the *neuO* gene was conserved, further indicating its importance to NMEC and thereby its potential role in receptor modification. Indeed, the data presented here revealed that deletion of *neuO* resulted in "hyper-susceptibility" to K1-specific bacteriophage, supporting the idea that NeuO prevents "super-infection" by modifying the bacteriophage receptor. Furthermore, it was revealed that deletion of *dsdC1* but not *dsdC2* caused a similar "hyper-susceptible" phenotype, suggesting that DsdC1 may be required for activation of NeuO. *O*-acetylation of LPS has been shown to effect phage adsorption, with acetylation protecting *E. coli* 4s from infection (reviewed in Prokhorov *et al.*, 2017). This potentially indicates that NeuO-mediated *O*-acetylation of the K1 capsule likely provides a similar

protection. Conversely, previous work by Vimr and colleagues suggested that NeuO had no effect against K1 bacteriophage-mediated lysis in archetypal NMEC strain RS218 (King, Steenbergen & Vimr., 2007) . However, NeuO-dependent survival was further evidenced by two distinct phenotypes that arose from the CE10 WT population. Through DNA sequencing, it was revealed that colonies that were protected from the K1 bacteriophage were NeuO "phase ON", whereas those that were rapidly killed were NeuO "phase OFF". The differences therefore in data reported, could be due to differences in the genetic makeup between CE10 and RS218.

Metabolic signals have been demonstrated to affect non-enzymatic acetylation in *E. coli*, with glucose, lactate, fructose, and xylose inducing acetylation (reviewed in Christensen *et al.*, 2019). Although D-serine did not appear to affect the transcriptional regulation of *neuO*, DsdC is the D-serine metabolism regulator, and therefore could integrate host metabolic signals with virulence. The results presented here indicated a DsdC1-specific regulatory mechanism that mediated transcription of *neuO*, which potentially indicates why CE10 carries two homologous TFs with seemingly redundant functionalities. However, the exact mechanism of DsdC1-specific regulation of *neuO* is yet to be solved. Preliminary data suggested that the phase variation of *neuO* affected transcription. Previous reports on *siaD*, a capsular polysialyltransferase in *N. meningitidis*, have revealed that the transcription of "phase OFF" *siaD* was prematurely terminated (Lavitola *et al.*, 1999); indicating that this mechanism is not unique to NMEC. Further, preliminary sequencing data has alluded to the hypothesis that DsdC1 affects the SSM of *neuO*. In the Δ*dsdC1* mutant, *neuO* was found to be in a constant "phase OFF" position. Although the regulatory mechanism at play has yet to be solved, it has been demonstrated in this thesis that DsdC1 affects the survival of CE10 against K1 bacteriophage-mediated lysis.

However, this is not the complete story of DsdC's role in *E. coli* gene regulation. The binding and transcriptomic data presented in Chapters 3 and 4 alluded to putative roles for DsdC in regulation of colibactin biosynthesis in CFT073, and LPS and capsule biosynthesis in CE10. Indeed, upon exposure to D-serine, it was revealed that DsdC modified the O-antigen length of LPS in CE10. Silver staining of LPS revealed two bands of increased intensity in the O-antigen side chain of

the ΔΔ*dsdC1/2* mutant compared to the WT. As alterations of the O-antigen can occur in response to environmental signals (Lerouge & Vanderleyden, 2002), it could be speculated that upon sensing D-serine, DsdC represses LPS biosynthesis genes, causing alterations in the exposed O-antigen as a defence mechanism for immune evasion. Furthermore, although no differences were observed in the assays performed to elucidate the role of DsdC in K1 capsule regulation; the possibility that there are further physiological effects mediated by DsdC, in regard to capsule formation, cannot be ruled out, with further experiments therefore required.

In an era where AMR is on the rise, it is crucial to understand the fundamental processes underpinning bacterial success. Indeed, through the widespread overuse and misuse of anti-microbials, multi-drug resistant bacteria are now endemic in certain regions of the world (reviewed in Christaki, Marcou & Tofarides, 2020). Approximately 700,000 people die annually from deaths attributed to AMR and it is estimated that by 2050 this figure could rise to 10 million, indicating it is a serious and imminent threat to the human population (reviewed in Christaki, Marcou & Tofarides, 2020). Due to the cytotoxic nature of antimicrobials, there is a strong selective pressure for bacteria to evolve to resist them, with genetic AMR arising from horizontally-acquired AMR genes or mutations within the genome (reviewed in Schrader, Vaubourgeix & Nathan, 2020). Anti-virulence (or anti-infective) strategies aim to supersede this inherent problem by using small molecules to block or interfere in bacterial virulence pathways (reviewed in Rasko & Sperandio, 2010). This, in theory, disables the bacteria from colonising the host or preventing toxin production, amongst other processes, while applying less evolutionary pressure to evolve resistance (reviewed in Rasko & Sperandio, 2010).

Further, anti-virulence strategies, in theory, do not compromise the host commensal microbiota in the same way that traditional antibiotics do, providing an extra layer of protection to the host (reviewed in Dickey, Cheung & Otto, 2017). Anti-bactericidal molecular inhibitors of LcrF, a TF involved in the regulation of *Yersinia pseudotuberculosis* T3SS, were revealed to reduce virulence *in vitro* and *in vivo* (Garrity-Ryan *et al*., 2010); indicating that the inhibition of virulence-associated transcriptional regulators are a viable novel

strategy for preventing infection. Due to the role that DsdC plays in the survival of CE10 against K1 bacteriophages, DsdC could potentially make an attractive target for anti-virulence strategies. A DsdC inhibitor, used in tandem with a K1-specific bacteriophage, could therefore present a novel pathogen-specific alternative to treating NMEC, without the use of antibiotics.

In conclusion, this work has elucidated the direct regulon of DsdC in two clinically relevant ExPEC strains, CFT073 and CE10. These data led to the conclusion that DsdC has been tailored in distinct *E. coli* strains to mediate strain-specific regulation. Further, a novel role of DsdC1 in protection against K1 bacteriophages has been elucidated, through the regulation of the *O*-acetyltransferase gene, *neuO*. Whilst *E. coli* cases are on the rise across Europe, these findings provide a greater understanding into the complex world of *E. coli* gene regulation, enabling for future utilisation of this knowledge for the potential development of new antibacterial strategies.

References

Abedon, S. T. (2021). Detection of bacteriophages: phage plaques. In: Harper D.R., Abedon S.T., Burrowes B.H., McConville M.L. (eds). *Bacteriophages* pp. 507–538. Springer International Publishing. doi.org/10.1007/978-3-319-41986-2_16.

Abreu, A. G., & Barbosa, A. S. (2017). How *Escherichia coli* circumvent complement-mediated killing. *Frontiers in Immunology*, *8*(452), pp. 1–6. doi.org/10.3389/fimmu.2017.00452.

Alteri, C. J., Smith, S. N., & Mobley, H. L. T. (2009). Fitness of *Escherichia coli* during urinary tract infection requires gluconeogenesis and the TCA cycle. *PLoS Pathogens*, *5*(5), pp. e1000448. doi.org/10.1371/journal.ppat.1000448.

Anderson, G. G., Palermo, J. J., Schilling, J. D., Roth, R., Heuser, J., & Hultgren, S. J. (2003). Intracellular bacterial biofilm-like pods in urinary tract infections. *Science*, *301*(5629), pp. 105–107. doi.org/10.1126/science.1084550.

Anfora, A. T., Haugen, B. J., Roesch, P., Redford, P., & Welch, R. A. (2007). Roles of serine accumulation and catabolism in the colonization of the murine urinary tract by *Escherichia coli* CFT073. *Infection and Immunity*, *75*(11), pp. 5298–5304. doi.org/10.1128/IAI.00652-07

Anfora, A. T., & Welch, R. A. (2006). DsdX is the second D-serine transporter in uropathogenic *Escherichia coli* clinical isolate CFT073. *Journal of Bacteriology*, *188*(18), pp. 6622–6628. doi.org/10.1128/JB.00634-06

Arndt, D., Grant, J. R., Marcu, A., Sajed, T., Pon, A., Liang, Y., & Wishart, D. S. (2016). PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Research*, *44*(W1), W16–W21. doi.org/10.1093/nar/gkw387

Arnster-Choder, O., & Wright, A. (1993). Transcriptional regulation of the *bgl* operon of *Escherichia coli* involves phosphotransferase system-mediated phosphorylation of a transcriptional antiterminator. *Journal of Cellular Biochemistry*, *51*(1), pp. 83–90. doi.org/doi: 10.1002/jcb.240510115

Azam, T. A., Iwata, A., Nishimura, A., Ueda, S., & Ishihama, A. (1999). Growth phase-dependent variation in protein composition of the *Escherichia coli* nuclei. *Journal*

of Bacteriology, *181*(20), pp. 6361–6370. doi: 10.1128/JB.181.20.6361-6370.1999

Babitzke, P. (1997). Regulation of tryptophan biosynthesis: Trp-ing the TRAP or how *Bacillus subtilis* reinvented the wheel. *Molecular Microbiology*, *26*(1), pp. 1–9. doi: 10.1046/j.1365-2958.1997.5541915.x.

Babu, M. M., & Teichmann, S. A. (2003a). Functional determinants of transcription factors in *Escherichia coli*: protein families and binding sites. Trends in Genetics, 19(2), pp. 75–79. doi: 10.1016/S0168-9525(02)00039-2.

Babu, M. M., & Teichmann, S. A. (2003b). Evolution of transcription factors and the gene regulatory network in *Escherichia coli*. *Nucleic Acids Research*, *31*(4), pp. 1234–1244. doi: 10.1093/nar/gkg210

Badger, J. L., & Kim, K. S. (1998). Environmental growth conditions influence the ability of *Escherichia coli* K1 To invade brain microvascular endothelial cells and confer serum resistance. *Infection and Immunity*, *66*(12), pp. 5692–5697. doi: 10.1128/IAI.66.12.5692-5697.1998.

Badger, J. L., Wass, C. A., & Kim, K. S. (2000). Identification of *Escherichia coli* K1 genes contributing to human brain microvascular endothelial cell invasion by differential fluorescence induction. *Molecular Microbiology*, *36*(1), pp. 174–182. doi: 10.1046/j.1365-2958.2000.01840.x.

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., Ren, J., Li, W. W., & Noble, W. S. (2009). MEME Suite: Tools for motif discovery and searching. *Nucleic Acids Research*, *37*, pp. W202–W208. doi.org/10.1093/nar/gkp335

Barbirz, S., Müller, J. J., Uetrecht, C., Clark, A. J., Heinemann, U., & Seckler, R. (2008). Crystal structure of *Escherichia coli* phage HK620 tailspike: Podoviral tailspike endoglycosidase modules are evolutionarily related. *Molecular Microbiology*, *69*(2), pp. 303–316. doi.org/10.1111/j.1365-2958.2008.06311.x

Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Schones, D. E., Wang, Z., Wei, G., Chepelev, I., & Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. *Cell*, *129*(4), pp. 823–837. doi.org/10.1016/j.cell.2007.05.009

Bateman, A., Martin, M. J., Orchard, S., Magrane, M., Agivetova, R., Ahmad, S., Alpi, E., Bowler-Barnett, E. H., Britto, R., Bursteinas, B., Bye-A-Jee, H., Coetzee, R.,

Cukura, A., Silva, A. Da, Denny, P., Dogan, T., Ebenezer, T. G., Fan, J., Castro, L. G., … Zhang, J. (2021). UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Research*, *49*(D1), pp. D480–D489. doi.org/10.1093/nar/gkaa1100

Bengoechea, J. A., Brandenburg, K., Arraiza, M. D., Seydel, U., Skurnik, M., & Moriyón, I. (2003). Pathogenic *Yersinia enterocolitica* strains increase the outer membrane permeability in response to environmental stimuli by modulating lipopolysaccharide fluidity and lipid A structure. *Infection and Immunity*, *71*(4), pp. 2014-2021. doi.org/10.1128/IAI.71.4.2014-2021.2003

Benz, R., Schmid, A., Wagner, W., & Goebel, W. (1989). Pore formation by the *Escherichia coli* hemolysin: Evidence for an association-dissociation equilibrium of the pore-forming aggregates. *Infection and Immunity*, *57*(3), pp. 887–895. doi: 10.1128/iai.57.3.887-895.1989

Bernabeu, M., Sánchez-Herrero, J. F., Huedo, P., Prieto, A., Hüttener, M., Rozas, J., & Juárez, A. (2019). Gene duplications in the *E. coli* genome: Common themes among pathotypes. *BMC Genomics*, *20*(313), pp. 1–11. doi.org/10.1186/s12864-019-5683-4

Bernard, E., Rolain, T., Courtin, P., Guillot, A., Langella, P., Hols, P., & Chapot-Chartier, M. P. (2011). Characterization of O-acetylation of N-acetylglucosamine: A novel structural variation of bacterial peptidoglycan. *Journal of Biological Chemistry*, *286*(27), pp. 23950–23958. doi.org/10.1074/jbc.M111.241414

Bertani, B., & Ruiz, N. (2018). Function and Biogenesis of Lipopolysaccharides. *EcoSal Plus*, *8*(1), pp. 1–33. doi.org/10.1128/ecosalplus.esp-0001-2018

Bervoets, I., & Charlier, D. (2019). Diversity, versatility and complexity of bacterial gene regulation mechanisms: Opportunities and drawbacks for applications in synthetic biology. *FEMS Microbiology Reviews*, *43*(3), pp. 304–339. doi.org/10.1093/femsre/fuz001

Bhakdi, S., & Tranum-Jensen, J. (1986). Membrane damage by pore-forming bacterial cytolysins. *Microbial Pathogenesis*, *1*, pp. 5–14. doi: 10.1016/0882-4010(86)90027-6.

Bobay, L. M., Touchon, M., & Rocha, E. P. C. (2014). Pervasive domestication of defective prophages by bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(33), pp. 12127–12132.

doi.org/10.1073/pnas.1405336111

Bolognesi, B., & Lehner, B. (2018). Protein Overexpression: Reaching the limit. *ELife*, *7*(e39804), pp. 1–3. doi.org/10.7554/eLife.34595

Bonacorsi, S., & Bingen, E. (2005). Molecular epidemiology of *Escherichia coli* causing neonatal meningitis. *International Journal of Medical Microbiology*, *295*(6–7), pp. 373–381. doi.org/10.1016/j.ijmm.2005.07.011

Bondy-Denomy, J., Qian, J., Westra, E. R., Buckling, A., Guttman, D. S., Davidson, A. R., & Maxwell, K. L. (2016). Prophages mediate defense against phage infection through diverse mechanisms. *ISME Journal*, *10*(12), pp. 2854–2866. doi.org/10.1038/ismej.2016.79

Bonocora, R. P., & Wade, J. T. (2015). ChIP-seq for genome-scale analysis of bacterial DNA-binding proteins. In I. Artsimovitch & T. Santangelo (Eds.), *Bacterial Transcriptional Control: Methods in molecular biology, 1276*, pp. 327–340. Humana Press. doi.org/10.1007/978-1-4939-2392-2_20

Bornstein-Forst, S. M., McFall, E., & Palchaudhuri, S. (1987). *In vivo* D-Serine deaminase transcription start sites in wild-type *Escherichia coli* and in dsdA promoter mutants. *Journal of Bacteriology*, *169*(3), pp. 1056–1060. doi.org/doi: 10.1128/jb.169.3.1056-1060.1987

Brauer, A. L., White, A. N., Learman, B. S., Johnson, A. O., & Armbruster, C. E. (2019). D-Serine degradation by *Proteus mirabilis* contributes to fitness during single-species and polymicrobial catheter-associated urinary tract infection. *MSphere*, *4*(1), pp. e00020-19. doi.org/10.1128/msphere.00020-19

Brenner, S. E., Hubbard, T., Murzin, A., & Chothia, C. (1995). Gene duplications in *H. influenza*. *Nature*, *378*(6553), pp. 140–140. doi: 10.1038/378140a0.

Browning, D. F., & Busby, S. J. W. (2004). The regulation of bacterial transcription initiation. *Nature Reviews Microbiology*, *2*(1), pp. 57–65. doi.org/10.1038/nrmicro787

Browning, D. F., & Busby, S. J. W. (2016). Local and global regulation of transcription initiation in bacteria. *Nature Reviews Microbiology*, *14*(10), pp. 638–650. doi.org/10.1038/nrmicro.2016.103

Browning, D. F., Butala, M., & Busby, S. J. W. (2019). Bacterial Transcription Factors: Regulation by Pick “N” Mix. *Journal of Molecular Biology*, *431*(20), pp. 4067–4077. doi.org/10.1016/j.jmb.2019.04.011

Browning, D. F., Grainger, D. C., & Busby, S. J. W. (2010). Effects of nucleoid-associated proteins on bacterial chromosome structure and gene expression. *Current Opinion in Microbiology*, *13*(6), pp. 773–780. doi.org/10.1016/j.mib.2010.09.013

Busby, S., & Ebright, R. H. (1994). Promoter structure, promoter recognition, and transcription activation in prokatyotes. *Cell*, *79*(5), pp. 743–746. doi: 10.1016/0092-8674(94)90063-9

Cameron, E. A., & Sperandio, V. (2015). Frenemies: signaling and nutritional integration in pathogen-microbiota-host interactions. *Cell Host and Microbe*, *18*(3), pp. 275–284. doi.org/10.1016/j.chom.2015.08.007

Canchaya, C., Proux, C., Fournous, G., Bruttin, A., & Brüssow, H. (2003). Prophage genomics. *Microbiology and Molecular Biology Reviews*, *67*(2), pp. 238–276. doi.org/10.1128/mmbr.67.2.238-276.2003

Casaburi, G., Duar, R. M., Brown, H., Mitchell, R. D., Kazi, S., Chew, S., Cagney, O., Flannery, R. L., Sylvester, K. G., Frese, S. A., Henrick, B. M., & Freeman, S. L. (2021). Metagenomic insights of the infant microbiome community structure and function across multiple sites in the United States. *Scientific Reports*, *11*(1472), pp. 1–12. doi.org/10.1038/s41598-020-80583-9

Cava, F., Lam, H., De Pedro, M. A., & Waldor, M. K. (2011). Emerging knowledge of regulatory roles of D-amino acids in bacteria. *Cellular and Molecular Life Sciences*, *68*, pp. 817–831. doi.org/10.1007/s00018-010-0571-8

Chaudhuri, R. R., & Pallen, M. J. (2006). xBASE, a collection of online databases for bacterial comparative genomics. *Nucleic Acids Research*, *34*, pp. D335–D337. doi.org/10.1093/nar/gkj140

Christaki, E., Marcou, M., & Tofarides, A. (2020). Antimicrobial resistance in bacteria: Mechanisms, evolution, and persistence. *Journal of Molecular Evolution*, *88*(1), pp. 26–40. doi.org/10.1007/s00239-019-09914-3

Christensen, D. G., Xie, X., Basisty, N., Byrnes, J., McSweeney, S., Schilling, B., & Wolfe, A. J. (2019). Post-translational protein acetylation: An elegant mechanism for bacteria to dynamically regulate metabolic functions. *Frontiers in Microbiology*, *10*(1604), pp. 1-22. doi.org/10.3389/fmicb.2019.01604

Cieslewicz, M., & Vimr, E. (1997). Reduced polysialic acid capsule expression in *Escherichia coli* K1 mutants with chromosomal defects in *kpsF*. *Molecular Microbiology*, *26*(2), pp. 237–249. doi: 10.1046/j.1365-2958.1997.5651942.x.

Colino, J., & Outschoorn, I. (1999). The form variation of the capsular polysaccharide K1 is not a critical virulence factor of *Escherichia coli* in a neonatal mouse model of infection. *Pathogenesis*, *27*(4), pp. 187–196. doi: 10.1006/mpat.1999.0291

Connell, H., Agace, W., Klemm, P., Schembri, M., Marild$, S., & Svanborg, C. (1996). Type 1 fimbrial expression enhances *Escherichia coli* virulence for the urinary tract. *Proceedings of the National Academy of Sciences USA*, *93*(18), pp. 9827–9832. doi: 10.1073/pnas.93.18.9827

Connolly, J. P. R., Finlay, B, B., & Roe, A. J. (2015). From ingestion to colonization: The influence of the host environment on regulation of the LEE encoded type III secretion system in enterohaemorrhagic *Escherichia coli*. *Frontiers in Microbiology*, *6*(568), pp. 1–15. doi.org/10.3389/fmicb.2015.00568

Connolly, J. P. R., Goldstone, R. J., Burgess, K., Cogdell, R. J., Beatson, S. A., Vollmer, W., Smith, D. G. E., & Roe, A. J. (2015). The host metabolite D-serine contributes to bacterial niche specificity through gene selection. *ISME Journal*, *9*(4), pp. 1039-1051. doi.org/10.1038/ismej.2014.242

Connolly, J. P. R., O'Boyle, N., Turner, N. C. A., Browning, D. F., & Roe, A. J. (2019). Distinct intraspecies virulence mechanisms regulated by a conserved transcription factor. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(39), pp. 19695–19704. doi.org/10.1073/pnas.1903461116

Connolly, J. P. R., Turner, N. C. A., Hallam, J. C., Rimbi, P. T., Flett, T., McCormack, M. J., Roe, A. J., & O'Boyle, N. (2021). d-Serine induces distinct transcriptomes in diverse *Escherichia coli* pathotypes. *Microbiology*, *167*(10), pp. 1-6. doi.org/10.1099/mic.0.001093

Cosloy, S. D., & McFall, E. (1973). Metabolism of D-Serine in *Escherichia coli* K-12:

Mechanism of growth inhibition. *Journal of Bacteriology*, *114*(2), pp. 685–694. doi: 10.1128/jb.114.2.685-694.1973.

Cross, A. S., Kim, K. S., Wright, D. C., Sadoff, J. C., & Gemski, P. (1986). Role of lipopolysaccharide and capsule in the serum resistance of bacteremic strains of *Escherichia coli*. *The Journal of Infectious Diseases*, *154*(3), pp. 497–503. doi: 10.1093/infdis/154.3.497.

Croxen, M. A., & Finlay, B. B. (2010). Molecular mechanisms of *Escherichia coli* pathogenicity. *Nature Reviews Microbiology*, *8*(1), pp. 26–38. doi.org/10.1038/nrmicro2265

D'haeseleer, P. (2006). What are DNA sequence motifs? *Nature Biotechnology*, *24*(4), pp. 423–425. doi:10.1038/nbt0406-423.

Dale, A. P., & Woodford, N. (2015). Extra-intestinal pathogenic *Escherichia coli* (ExPEC): Disease, carriage and clones. *Journal of Infection*, *71*(6), pp. 615–626. doi.org/10.1016/j.jinf.2015.09.009

Dame, R. T. (2005). The role of nucleoid-associated proteins in the organization and compaction of bacterial chromatin. *Molecular Microbiology, 56*(4), pp. 858–870. doi.org/10.1111/j.1365-2958.2005.04598.x

Dame, R. T., Wyman, C., & Goosen, N. (2000). H-NS mediated compaction of DNA visualised by atomic force microscopy. *Nucleic Acids Research*, *28*(18), pp. 3504–3510. doi: 10.1093/nar/28.18.3504.

Datsenko, K. A., & Wanner, B. L. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proceedings of the National Academy of Sciences*, *97*(12), pp. 6640–6645. doi.org/10.1073/pnas.120163297

Decker, E. M., Klein, C., Schwindt, D., & Von Ohle, C. (2014). Metabolic activity of *Streptococcus mutans* biofilms and gene expression during exposure to xylitol and sucrose. *International Journal of Oral Science*, *6*, pp. 195–204. doi.org/10.1038/ijos.2014.38

Denamur, E., Clermont, O., Bonacorsi, S., & Gordon, D. (2021). The population genetics of pathogenic *Escherichia coli*. *Nature Reviews Microbiology*, *19*(1), pp. 37–54. doi.org/10.1038/s41579-020-0416-x

Deszo, E. L., Steenbergen, S. M., Freedberg, D. I., Vimr, E. R., & Robbins, J. B. (2005). *Escherichia coli* K1 polysialic acid O-acetyltransferase gene, *neuO*, and the mechanism of capsule form variation involving a mobile contingency locus. *Proceedings of the National Academy of Sciences*, *102*(15), pp. 5564–5569. doi: 10.1073/pnas.0407428102.

Dhakal, B. K., & Mulvey, M. A. (2012). The UPEC pore-forming toxin α-hemolysin triggers proteolysis of host proteins to disrupt cell adhesion, inflammatory, and survival pathways. *Cell Host and Microbe*, *11*(1), pp. 58–69. doi.org/10.1016/j.chom.2011.12.003

Dickey, S. W., Cheung, G. Y. C., & Otto, M. (2017). Different drugs for bad bugs: Antivirulence strategies in the age of antibiotic resistance. *Nature Reviews Drug Discovery*, *16*(7), pp. 457–471. doi.org/10.1038/nrd.2017.23

Dillon, S. C., Cameron, A. D. S., Hokamp, K., Lucchini, S., Hinton, J. C. D., & Dorman, C. J. (2010). Genome-wide analysis of the H-NS and Sfh regulatory networks in *Salmonella* Typhimurium identifies a plasmid-encoded transcription silencing mechanism. *Molecular Microbiology*, *76*(5), pp. 1250–1265. doi.org/10.1111/j.1365-2958.2010.07173.x

Dillon, S. C., & Dorman, C. J. (2010). Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nature Reviews Microbiology*, *8*(3), pp. 185–195. doi.org/10.1038/nrmicro2261

Dillon, S. C., Espinosa, E., Hokamp, K., Ussery, D. W., Casadesús, J., & Dorman, C. J. (2012). LeuO is a global regulator of gene expression in *Salmonella enterica* serovar Typhimurium. *Molecular Microbiology*, *85*(6), pp. 1072–1089. doi.org/10.1111/j.1365-2958.2012.08162.x

Dorman, C. J. (2007). H-NS, the genome sentinel. *Nature Reviews Microbiology*, *5*(2), pp. 157–161. doi:10.1038/nrmicro1598.

Dorman, C. J., Schumacher, M. A., Bush, M. J., Brennan, R. G., & Buttner, M. J. (2020). When is a transcription factor a NAP? *Current Opinion in Microbiology*, *55*, pp. 26–33. doi.org/10.1016/j.mib.2020.01.019

Dunkelberger, J. R., & Song, W. C. (2010). Complement and its role in innate and adaptive immune responses. *Cell Research*, *20*(1), pp. 34–50.

doi.org/10.1038/cr.2009.139

Dunn, T. M., Hahn, S., Ogdent, S., & Schleif, R. F. (1984). An operator at-280 base pairs that is required for repression of *araBAD* operon promoter: Addition of DNA helical turns between the operator and promoter cyclically hinders repression. *Proceedings of the National Academy of Sciences, 81*, pp. 5017–5020. doi: 10.1073/pnas.81.16.5017.

Ebbensgaard, A., Mordhorst, H., Aarestrup, F. M., & Hansen, E. B. (2018). The role of outer membrane proteins and lipopolysaccharides for the sensitivity of *Escherichia coli* to antimicrobial peptides. *Frontiers in Microbiology*, *9*(2153), pp. 1–13. doi.org/10.3389/fmicb.2018.02153

Essler, M., Linder, S., Schell, B., Hüfner, K., Wiedemann, A., Randhahn, K., Staddon, J. M., & Aepfelbacher, M. (2003). Cytotoxic necrotizing factor 1 of *Escherichia coli* stimulates Rho/Rho-kinase-dependent myosin light-chain phosphorylation without inactivating myosin light-chain phosphatase in endothelial cells. *Infection and Immunity*, *71*(9), pp. 5188–5193. doi.org/10.1128/IAI.71.9.5188-5193.2003

Ezezika, O. C., Haddad, S., Neidle, E. L., & Momany, C. (2007). Oligomerization of BenM, a LysR-type transcriptional regulator: Structural basis for the aggregation of proteins in this family. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications, 63*(5), pp. 361–368. doi.org/10.1107/S1744309107019185

Fabbri, A., Travaglione, S., & Fiorentini, C. (2010). *Escherichia coli* cytotoxic necrotizing factor 1 (CNF1): Toxin biology, *in vivo* applications and therapeutic potential. *Toxins, 2*, pp. 283–296. doi.org/10.3390/toxins2020282

Faïs, T., Delmas, J., Barnich, N., Bonnet, R., & Dalmasso, G. (2018). Colibactin: More than a new bacterial toxin. *Toxins*, *10*(151), pp. 1-16. doi.org/10.3390/toxins10040151

Fang, Z., & Cui, X. (2011). Design and validation issues in RNA-seq experiments. *Briefings in Bioinformatics*, *12*(3), pp. 280–287. doi.org/10.1093/bib/bbr004

Feklístov, A., Sharon, B. D., Darst, S. A., & Gross, C. A. (2014). Bacterial sigma factors: A historical, structural, and genomic perspective. *Annual Review of Microbiology*, *68*, pp. 357–376. doi.org/10.1146/annurev-micro-092412-155737

Felmlee, T., Pellett, S., & Welch, R. A. (1985). Nucleotide sequence of an *Escherichia coli* chromosomal hemolysin. *Journal of Bacteriology*, *163*(1), pp. 94–105. doi: 10.1128/jb.163.1.94-105.1985.

Flores-Mireles, A. L., Walker, J. N., Caparon, M., & Hultgren, S. J. (2015). Urinary tract infections: Epidemiology, mechanisms of infection and treatment options. *Nature Reviews Microbiology*, *13*(5), pp. 269–284. doi.org/10.1038/nrmicro3432

Fomsgaard, A., Freudenberg, M. A., & Galanos, C. (1990). Modification of the silver staining technique to detect lipopolysaccharide in polyacrylamide gels. *Journal of Clinical Microbiology*, *28*(12), pp. 2627–2631. doi: 10.1128/jcm.28.12.2627-2631.1990.

Fox, S., Goswami, C., Holden, M., Connolly, J. P. R., Mordue, J., O'Boyle, N., Roe, A., Connor, M., Leanord, A., & Evans, T. J. (2020). A highly conserved complete accessory *Escherichia coli* type III secretion system 2 is widespread in bloodstream isolates of the ST69 lineage. *Scientific Reports*, *10*(4135), pp. 1–11. doi.org/10.1038/s41598-020-61026-x

Frasa, H., Procee, J., Torensma, R., Verbruggen, A., Algra, A., Rozenberg-Arska, M., Kraaijeveld, K., & Verhoef Eijkman-Winkler, J. (1993). *Escherichia coli* in bacteremia: O-acetylated K1 strains appear to be more virulent than non-O-acetylated K1 strains. *Journal of Clinical Microbiology*, *31*(12), pp. 3174–3178. doi: 10.1128/jcm.31.12.3174-3178.1993.

Friedman, M. (2010). Origin, microbiology, nutrition, and pharmacology of D-amino acids. *Chemistry and Biodiversity*, *7*(6), pp. 1491–1530. doi: 10.1002/cbdv.200900225.

Fusco, P. C., Farley, E. K., Huang, C. H., Moore, S., & Michon, F. (2007). Protective meningococcal capsular polysaccharide epitopes and the role of O acetylation. *Clinical and Vaccine Immunology*, *14*(5), pp. 577–584. doi.org/10.1128/CVI.00009-07

Galagan, J., Lyubetskaya, A., & Gomes, A. (2012). ChIP-Seq and the Complexity of Bacterial Transcriptional Regulation. *Systems Biology: Current topics in microbiology and immunology*. (M. Katza (ed.); Vol. 363). pp. 43-68. Springer. doi: 10.1007/82_2012_257.

Galas, D. J., & Schmitz, A. (1978). DNAse footprinting: a simple method for the detection of protein-DNA binding specificity. *Nucleic Acids Research*, *5*(9), pp. 3157–3170. doi:10.1093/nar/5.9.3157.

Gallet, R., Kannoly, S., & Wang, I. N. (2011). Effects of bacteriophage traits on plaque formation. *BMC Microbiology*, *11*(181), pp. 1–16. doi.org/10.1186/1471-2180-11-181

Gao, C., Garren, M., Penn, K., Fernandez, V. I., Seymour, J. R., Thompson, J. R., Raina, J. B., & Stocker, R. (2021). Coral mucus rapidly induces chemokinesis and genome-wide transcriptional shifts toward early pathogenesis in a bacterial coral pathogen. *ISME Journal*, *15*, pp. 1–15. doi.org/10.1038/s41396-021-01024-7

Garcia, E. C., Brumbaugh, A. R., & Mobley, H. L. T. (2011). Redundancy and specificity of *Escherichia coli* iron acquisition systems during urinary tract infection. *Infection and Immunity*, *79*(3), pp. 1225–1235. doi.org/10.1128/IAI.01222-10

Garrity-Ryan, L. K., Kim, O. K., Balada-Llasat, J. M., Bartlett, V. J., Verma, A. K., Fisher, M. L., Castillo, C., Songsungthong, W., Tanaka, S. K., Levy, S. B., Mecsas, J., & Alekshun, M. N. (2010). Small molecule inhibitors of LcrF, a *Yersinia pseudotuberculosis* transcription factor, attenuate virulence and limit infection in a murine pneumonia model. *Infection and Immunity*, *78*(11), pp. 4683–4690. doi.org/10.1128/IAI.01305-09

Genchi, G. (2017). An overview on D-amino acids. *Amino Acids*, *49*, pp. 1521–1533. doi.org/10.1007/s00726-017-2459-5

Geno, K. A., Gilbert, G. L., Song, J. Y., Skovsted, I. C., Klugman, K. P., Jones, C., Konradsen, H. B., & Nahm, M. H. (2015). Pneumococcal capsules and their types: Past, present, and future. *Clinical Microbiology Reviews*, *28*(3), pp. 871–899. doi.org/10.1128/CMR.00024-15

Ghosh, T., Bose, D., & Zhang, X. (2010). Mechanisms for activating bacterial RNA polymerase. *FEMS Microbiology Reviews*, *34*(5), pp. 611–627. doi.org/10.1111/j.1574-6976.2010.00239.x

Glynn, A. A., & Howard, C. J. (1970). The sensitivity to complement of strains of *Escherichia coli* related to their K antigens. *Immunology*, *18*(3), pp. 331–346. PMID: 4986073.

Goethals, K., Van Montagu, M., & Holsters, M. (1992). Conserved motifs in a divergent nod box of *Azorhizobium caulinodans* ORS571 reveal a common structure in promoters regulated by LysR-type proteins. *Proceedings of the National Academy of Sciences USA*, *89*(5), pp. 1646–1650. doi: 10.1073/pnas.89.5.1646.

Goh, K. G. K., Phan, M. D., Forde, B. M., Chong, T. M., Yin, W. F., Chan, K. G., Ulett, G. C., Sweet, M. J., Beatson, S. A., & Schembri, M. A. (2017). Genome-wide discovery of genes required for capsule production by uropathogenic *Escherichia coli*. *MBio*, *8*(5), pp. e01558-17. doi.org/10.1128/mBio.01558-17

Grainger, D. C., Goldberg, M. D., Lee, D. J., & Busby, S. J. W. (2008). Selective repression by Fis and H-NS at the *Escherichia coli dps* promoter. *Molecular Microbiology*, *68*(6), pp. 1366–1377. doi.org/10.1111/j.1365-2958.2008.06253.x

Grainger, D. C., Hurd, D., Goldberg, M. D., & Busby, S. J. W. (2006). Association of nucleoid proteins with coding and non-coding segments of the *Escherichia coli* genome. *Nucleic Acids Research*, *34*(16), pp. 4642–4652. doi.org/10.1093/nar/gkl542

Hacker, J., & Blum-Oehler, G. (2007). In appreciation of Theodor Escherich. *Nature Reviews Microbiology*, *5*, pp. 902. doi.org/10.1038/nrmicro1810

Hammerschmidt, S., Muller, A., Sillmann, H., Muhlenhoff, M., Borrow, R., Fox, A., Van Putten, J., Zollinger, W. D., Gerardy-Schahn, R., & Frosch, M. (1996). Capsule phase variation in *Neisseria meningitidis* serogroup B by slipped-strand mispairing in the polysialyltransferase gene (*siaD*): correlation with bacterial invasion and the outbreak of meningococcal disease. *Molecular Microbiology*, *20*(6), pp. 1211–1220. doi: 10.1111/j.1365-2958.1996.tb02641.x.

Haraoka, M., Hang, L., Frendéus, B., Godaly, G., Burdick, M., Strieter, R., & Svanborg, C. (1999). Neutrophil recruitment and resistance to urinary tract infection. *The Journal of Infectious Diseases*, *180*(4), pp. 1220–1229. doi: 10.1086/315006.

Haugen, B. J., Pellett, S., Redford, P., Hamilton, H. L., Roesch, P. L., & Welch, R. A. (2007). *In vivo* gene expression analysis identifies genes required for enhanced colonization of the mouse urinary tract by uropathogenic *Escherichia coli* strain CFT073 *dsdA*. *Infection and Immunity*, *75*(1), pp. 278–289. doi.org/10.1128/IAI.01319-06

Heincz, M. C., Bornstein, S. M., & McFall, E. (1984). Purification and characterization of D-Serine deaminase activator protein. *Journal of Bacteriology*, *160*(1), pp. 42–49. doi: 10.1128/jb.160.1.42-49.1984.

Heincz, M. C., & McFall, E. (1978). Role of the *dsdC* activator in regulation of D-Serine deaminase synthesis. *Journal of Bacteriology*, *136*(1), pp. 96–103. doi: 10.1128/jb.136.1.96-103.1978.

Heinrichs, D. E., Yethon, J. A., Amor, P. A., & Whitfield, C. (1998). The assembly system for the outer core portion of R1- and R4-type lipopolysaccharides of *Escherichia coli*: the R1 core-specific b-glucosyltransferase provides a novel attachment site for O-polysaccharides. *Journal of Biological Chemistry*, *273*(45), pp. 29497–29505. doi:10.1074/jbc.273.45.29497.

Heldwein, E. E. Z., & Brennan, R. G. (2001). Crystal structure of the transcription activator BmrR bound to DNA and a drug. *Nature*, *409*(6818), pp. 378–382. doi: 10.1038/35053138.

Hellman, L. M., & Fried, M. G. (2007). Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. *Nature Protocols*, *2*(8), pp. 1849–1861. doi.org/10.1038/nprot.2007.249

Henderson, I. R., Owen, P., & Nataro, J. P. (1999). Molecular switches - the ON and OFF of bacterial phase variation. *Molecular Microbiology*, *33*(5), pp. 919–932. doi: 10.1046/j.1365-2958.1999.01555.x.

Henikoff, S., Haughno, G. W., Calvot, J. M., & Wallace, J. C. (1988). A large family of bacterial activator proteins. *Proceedings of the National Academy of Sciences USA*, *85*(18), pp. 6602–6606. doi: 10.1073/pnas.85.18.6602.

Henkin, T. M. (2000). Transcription termination control in bacteria. *Current Opinion in Microbiology*, *3*(2), pp. 149–153. doi: 10.1016/s1369-5274(00)00067-9.

Hibbing, M. E., Fuqua, C., Parsek, M. R., & Peterson, S. B. (2010). Bacterial competition: Surviving and thriving in the microbial jungle. *Nature Reviews Microbiology*, *8*(1), pp. 15–25. doi.org/10.1038/nrmicro2259

Ho-Sui, S. J., Fedynak, A., Hsiao, W. W. L., Langille, M. G. I., & Brinkman, F. S. L. (2009). The association of virulence factors with genomic Islands. *PLoS ONE*, *4*(12),

pp. 1–11. doi.org/10.1371/journal.pone.0008094

Holden, N. S., & Tacon, C. E. (2011). Principles and problems of the electrophoretic mobility shift assay. *Journal of Pharmacological and Toxicological Methods*, *63*(1), pp. 7–14. doi.org/10.1016/j.vascn.2010.03.002

Hooton, T. M. (2012). Uncomplicated urinary tract infection. *The New England Journal of Medicine*, *366*(11), pp. 1028–1037. doi.org/10.1056/NEJMcp1104429

Hryckowian, A. J., Baisa, G. A., Schwartz, K. J., & Welch, R. A. (2015). DsdA does not affect colonization of the murine urinary tract by *Escherichia coli* CFT073. *PLoS ONE*, *10*(9), pp. e0138121. doi.org/10.1371/journal.pone.0138121

Huang, S-H., Chen, Y.-H., Fu, Q. I., Stins, M., Wang, Y., Wass, C., Kwang, A., & Kim, S. (1999). Identification and characterization of an *Escherichia coli* invasion gene locus, *ibeB*, required for penetration of brain microvascular endothelial cells. *Infection and Immunity*, *67*(5), pp. 2103–2109. doi: 10.1128/IAI.67.5.2103-2109.1999.

Huang, S-H., Wass, C., Fu, Q. I., Prasadarao, N. V, Stins, M., Kwang, A., & Kim, S. (1995). *Escherichia coli* invasion of brain microvascular endothelial cells *in vitro* and *in vivo*: molecular cloning and characterization of invasion gene *ibe10*. *Infection and Immunity*, *63*(11), pp. 4470–4475. doi: 10.1128/iai.63.11.4470-4475.1995.

Ishihama, A. (2000). Functional modulation of *Escherichia coli* RNA polymerase. *Annual Review of Microbiology*, *54*, pp. 499–518. doi: 10.1146/annurev.micro.54.1.499.

Iwata, Y., Sakai, N., Yoneda, I., Senda, Y., Sakai-Takemori, Y., Oshima, M., Nakagawa-Yoneda, S., Ogura, H., Sato, K., Minami, T., Kitajima, S., Toyama, T., Yamamura, Y., Miyagawa, T., Hara, A., Shimizu, M., Furuichi, K., Matsushima, K., & Wada, T. (2021). D-Serine inhibits the attachment and biofilm formation of methicillin-resistant *Staphylococcus aureus*. *Biochemical and Biophysical Research Communications*, *537*, pp. 50–56. doi.org/10.1016/j.bbrc.2020.12.078.

Jacob, F., & Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology*, *3*, pp. 318–356. doi.org/10.1016/S0022-2836(61)80072-7.

Jain, D., Baldi, S., Zabel, A., Straub, T., & Becker, P. B. (2015). Active promoters give rise to false positive “Phantom Peaks” in ChIP-seq experiments. *Nucleic Acids Research*, *43*(14), pp. 6959–6968. doi.org/10.1093/nar/gkv637

Jamieson, J. J., Searson, P. C., & Gerecht, S. (2017). Engineering the human blood-brain barrier *in vitro*. *Journal of Biological Engineering*, *11*(37), pp. 1-11. doi.org/10.1186/s13036-017-0076-1

Jann, K., & Jann, B. (1992). Capsules of *Escherichia coli*, expression and biological significance. *Canadian Journal of Microbiology*, *38*(7), pp. 705–710. doi: 10.1139/m92-116.

Jia, J., King, J. E., Goldrick, M. C., Aldawood, E., & Roberts, I. S. (2017). Three tandem promoters, together with IHF, regulate growth phase dependent expression of the *Escherichia coli kps* capsule gene cluster. *Scientific Reports*, *7*(1), pp. 1-11. doi.org/10.1038/s41598-017-17891-0

Johnson, David, S., Mortazavi, A., Myers, Richard, M., & Wold, B. (2007). Genome-wide mapping of *in vivo* protein-DNA interactions. *Science*, *316*(5830), pp. 1497–1502. doi.org/10.1126/science.1142265

Johnson, J. R., & Russo, T. A. (2002). Extraintestinal pathogenic *Escherichia coli*: “The other bad *E coli*.” *Journal of Laboratory and Clinical Medicine*, *139*(3), pp. 155–162. doi.org/10.1067/mlc.2002.121550

Juhas, M. (2015). Horizontal gene transfer in human pathogens. *Critical Reviews in Microbiology*, *41*(1), pp. 101–108. doi.org/10.3109/1040841X.2013.804031

Jurczak-Kurek, A., Gasior, T., Nejman-Faleńczyk, B., Bloch, S., Dydecka, A., Topka, G., Necel, A., Jakubowska-Deredas, M., Narajczyk, M., Richert, M., Mieszkowska, A., Wróbel, B., Węgrzyn, G., & Węgrzyn, A. (2016). Biodiversity of bacteriophages: Morphological and biological properties of a large group of phages isolated from urban sewage. *Scientific Reports*, *6*(34338), pp. 1-17. https://doi.org/10.1038/srep34338

Kahramanoglou, C., Seshasayee, A. S. N., Prieto, A. I., Ibberson, D., Schmidt, S., Zimmermann, J., Benes, V., Fraser, G. M., & Luscombe, N. M. (2011). Direct and indirect effects of H-NS and Fis on global gene expression control in *Escherichia coli*. *Nucleic Acids Research*, *39*(6), pp. 2073–2091. doi.org/10.1093/nar/gkq934

Kaper, J. B., Nataro, J. P., & Mobley, H. L. T. (2004). Pathogenic *Escherichia coli*. *Nature Reviews Microbiology*, *2*, pp. 123–140. doi.org/10.1038/nrmicro818

Kapust, R. B., & Waugh, D. S. (1999). *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Science*, *8*(8), pp. 1668–1674. doi: 10.1110/ps.8.8.1668.

Khalessi, N., & Afsharkas, L. (2014). Neonatal meningitis: Risk factors, causes, and neurologic complications. *Iran J Child Neurol*, *8*(4), pp. 46–50. PCMID: PMC4307368.

Khan, N. A., Kim, Y., Shin, S., & Kim, K. S. (2007). FimH-mediated *Escherichia coli* K1 invasion of human brain microvascular endothelial cells. *Cellular Microbiology*, *9*(1), pp. 169–178. doi.org/10.1111/j.1462-5822.2006.00779.x

Khan, N. A., Wang, Y., Kim, K. J., Chung, J. W., Wass, C. A., & Kim, K. S. (2002). Cytotoxic necrotizing factor-1 contributes to *Escherichia coli* K1 invasion of the central nervous system. *Journal of Biological Chemistry*, *277*(18), pp. 15607–15612. doi.org/10.1074/jbc.M112224200

Kim, K. J., Chung, J. W., & Kim, K. S. (2005). 67-kDa laminin receptor promotes internalization of cytotoxic necrotizing factor 1-expressing *Escherichia coli* K1 into human brain microvascular endothelial cells. *Journal of Biological Chemistry*, *280*(2), pp. 1360–1368. doi.org/10.1074/jbc.M410176200

Kim, K. S. (2012). Current concepts on the pathogenesis of *Escherichia coli* meningitis: Implications for therapy and prevention. *Current Opinion in Infectious Diseases*, *25*(3), pp. 273–278. doi.org/10.1097/QCO.0b013e3283521eb0

Kim, K. S. (2000). *E. coli* invasion of brain microvascular endothelial cells as a pathogenetic basis of meningitis. In T. Oelschlaeger & J. Hacker (Eds.), *Bacterial Invasion into Eukaryotic Cells. Subcellular Biochemistry. 33*, pp. 47–59. doi: 10.1007/978-1-4757-4580-1_3.

Kim, K. S, Itabashi, H., Gemski, P., Sadoff, J., Warren, R. L., & Cross, A. S. (1992). The K1 capsule is the critical determinant in the development of *Escherichia coli* meningitis in the rat. *Journal of Clinical Investigation*, *90*(3), pp. 897–905. doi.org/10.1172/JCI115965

King, C. H., Desai, H., Sylvetsky, A. C., LoTempio, J., Ayanyan, S., Carrie, J., Crandall,

K. A., Fochtman, B. C., Gasparyan, L., Gulzar, N., Howell, P., Issa, N., Krampis, K., Mishra, L., Morizono, H., Pisegna, J. R., Rao, S., Ren, Y., Simonyan, V., … Mazumder, R. (2019). Baseline human gut microbiota profile in healthy people and standard reporting template. *PLoS ONE*, *14*(9), pp. e0206484. doi.org/10.1371/journal.pone.0206484

King, M. R., Steenbergen, S. M., & Vimr, E. R. (2007). Going for baroque at the *Escherichia coli* K1 cell surface. *Trends in Microbiology*, *15*(5), pp. 196–202. doi.org/10.1016/j.tim.2007.03.006

King, M. R., Vimr, R. P., Steenbergen, S. M., Spanjaard, L., Plunkett, G., Blattner, F. R., & Vimr, E. R. (2007). *Escherichia coli* K1-specific bacteriophage CUS-3 distribution and function in phase-variable capsular polysialic acid O acetylation. *Journal of Bacteriology*, *189*(17), pp. 6447–6456. doi.org/10.1128/JB.00657-07

Kleinman, C. L., Sycz, G., Bonomi, H. R., Rodriguez, R. M., Zorreguieta, A., & Sieira, R. (2017). ChIP-seq analysis of the LuxR-type regulator VjbR reveals novel insights into the *Brucella* virulence gene expression network. *Nucleic Acids Research*, *45*(10), pp. 5757–5769. doi.org/10.1093/nar/gkx165

Koide, T., Reiss, D. J., Bare, J. C., Pang, W. L., Facciotti, M. T., Schmid, A. K., Pan, M., Marzolf, B., Van, P. T., Lo, F. Y., Pratap, A., Deutsch, E. W., Peterson, A., Martin, D., & Baliga, N. S. (2009). Prevalence of transcription promoters within archaeal operons and coding sequences. *Molecular Systems Biology*, *5*(285), pp. 1–16. doi.org/10.1038/msb.2009.42

Kolodkin-Gal, I., Romero, D., Cao, S., Clardy, J., Kolter, R., & Losick, R. (2010). D-amino acids trigger biofilm disassembly. *Science*, *328*(5978), pp. 627–629. https://doi.org/10.1126/science.1187113

Korotkov, K. V., Johnson, T. L., Jobling, M. G., Pruneda, J., Pardon, E., Héroux, A., Turley, S., Steyaert, J., Holmes, R. K., Sandkvist, M., & Hol, W. G. J. (2011). Structural and functional studies on the interaction of GspC and GspD in the type ii secretion system. *PLoS Pathogens*, *7*(9), pp. e1002228. doi.org/10.1371/journal.ppat.1002228

Korte-Berwanger, M., Sakinc, T., Kline, K., Nielsen, H. V., Hultgren, S., & Gatermann, S. G. (2013). Significance of the D-serine-deaminase and D-serine metabolism of *Staphylococcus saprophyticus* for virulence. *Infection and Immunity*, *81*(12), pp.

4525–4533. doi.org/10.1128/IAI.00599-13

Ku, L. C., Boggess, K. A., & Cohen-Wolkowiez, M. (2015). Bacterial meningitis in infants. *Clinics in Perinatology*, *42*(1), pp. 29–45. doi.org/10.1016/j.clp.2014.10.004

Kulkarni, R., Dhakal, B. K., Slechta, E. S., Kurtz, Z., Mulvey, M. A., & Thanassi, D. G. (2009). Roles of putative type II secretion and type IV pilus systems in the virulence of uropathogenic *Escherichia coli*. *PLoS ONE*, *4*(3), pp. e4752. doi.org/10.1371/journal.pone.0004752

Labrie, S. J., Samson, J. E., & Moineau, S. (2010). Bacteriophage resistance mechanisms. *Nature Reviews Microbiology*, *8*(5), pp. 317–327. doi.org/10.1038/nrmicro2315

Lam, H., Oh, D-C., Cava, F., Takacs, C. N., Clardy, J., de Pedro, M. A., & Waldor, M. K. (2009). D-amino acids govern stationary phase cell wall remodeling in bacteria. *Science*, *325*(5947), pp. 1552–1555. doi.org/10.1098/rsif.2008.0435

Lavitola, A., Bucci, C., Salvatore, P., Maresca, G., Bruni, C. B., & Alifano, P. (1999). Intracistronic transcription termination in polysialyltransferase gene (*siaD*) affects phase variation in *Neisseria meningitidis*. *Molecular Microbiology*, *33*(1), pp. 119–127. doi: 10.1046/j.1365-2958.1999.01454.x.

Lemichez, E., Flatau, G., Bruzzone, M., Boquet, P., & Gauthier, M. (1997). Molecular localization of the *Escherichia coli* cytotoxic necrotizing factor CNF1 cell-binding and catalytic domains. *Molecular Microbiology*, *24*(5), pp. 1061–1070. doi: 10.1046/j.1365-2958.1997.4151781.x.

Lerouge, I., & Vanderleyden, J. (2002). O-antigen structural variation: mechanisms and possible roles in animal/plant–microbe interactions. *FEMS Microbiology Reviews*, *26*(1), pp. 17–47. doi.org/10.1111/j.1574-6976.2002.tb00597.x

Leying, H., Suerbaum, S., Kroll, H-P., Stahl, D. and Opferkuch, W. (1990). The capsular polysaccharide is a major determinant of serum resistance in K-1-positive blood culture isolates of *Escherichia coli*. *Infection and Immunity*, *58*(1), 222–227. doi: 10.1128/iai.58.1.222-227.1990.

Li, H., Limenitakis, J. P., Fuhrer, T., Geuking, M. B., Lawson, M. A., Wyss, M., Brugiroux, S., Keller, I., Macpherson, J. A., Rupp, S., Stolp, B., Stein, J. V.,

Stecher, B., Sauer, U., McCoy, K. D., & Macpherson, A. J. (2015). The outer mucus layer hosts a distinct intestinal microbial niche. *Nature Communications*, *6*(8292), pp. 1–13. doi.org/10.1038/ncomms9292

Li, K., Wu, G., Liao, Y., Zeng, Q., Wang, H., & Liu, F. (2020). RpoN1 and RpoN2 play different regulatory roles in virulence traits, flagellar biosynthesis, and basal metabolism in *Xanthomonas campestris*. *Molecular Plant Pathology*, *21*(7), pp. 907–922. doi.org/10.1111/mpp.12938

Li, Z., Quan, G., Jiang, X., Yang, Y., Ding, X., Zhang, D., Wang, X., Hardwidge, P. R., Ren, W., & Zhu, G. (2018). Effects of metabolites derived from gut microbiota and hosts on pathogens. *Frontiers in Cellular and Infection Microbiology*, *8*(314), pp. 1-12. doi.org/10.3389/fcimb.2018.00314

Lin, S-Y., & Riggs, A. D. (1975). The general affinity of *lac* repressor for *E. coli* DNA : implications for gene regulation in procaryotes and eucaryotes. *Cell*, *4*(2), pp. 107–111. doi: 10.1016/0092-8674(75)90116-6.

Lindquist, S., Lindberg, F., & Normark, S. (1989). Binding of the *Citrobacter freundii* AmpR regulator to a single DNA site provides both autoregulation and activation of the inducible *ampC* B-Lactamase gene. *Journal of Bacteriology*, *171*(7), pp. 3746–3753. doi: 10.1128/jb.171.7.3746-3753.1989.

Liu, B., Furevi, A., Perepelov, A. V., Guo, X., Cao, H., Wang, Q., Reeves, P. R., Knirel, Y. A., Wang, L., & Widmalm, G. (2020). Structure and genetics of *Escherichia coli* O antigens. *FEMS Microbiology Reviews*, *44*(6), pp. 655–683. doi.org/10.1093/femsre/fuz028

Liu, C., Zheng, H., Yang, M., Xu, Z., Wang, X., Wei, L., Tang, B., Liu, F., Zhang, Y., Ding, Y., Tang, X., Wu, B., Johnson, T. J., Chen, H., & Tan, C. (2015). Genome analysis and *in vivo* virulence of porcine extraintestinal pathogenic *Escherichia coli* strain PCN033. *BMC Genomics*, *16*(717) pp. 1-18. doi.org/10.1186/s12864-015-1890-9

Lonetto, M., Gribskov, M., & Gross, C. A. (1992). The sigma factor 70 family: sequence conservation and evolutionary relationships. *Journal of Bacteriology*, *174*(12), pp. 3843–3849. doi: 10.1128/jb.174.12.3843-3849.1992.

Lopez, M. S., Tan, I. S., Yan, D., Kang, J., McCreary, M., Modrusan, Z., Austin, C. D.,

Xu, M., & Brown, E. J. (2017). Host-derived fatty acids activate type VII secretion in *Staphylococcus aureus*. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(42), pp. 11223–11228. doi.org/10.1073/pnas.1700627114

Lu, S., Jin, D., Wu, S., Yang, J., Lan, R., Bai, X., Liu, S., Meng, Q., Yuan, X., Zhou, J., Pu, J., Chen, Q., Dai, H., Hu, Y., Xiong, Y., Ye, C., & Xu, J. (2016). Insights into the evolution of pathogenicity of *Escherichia coli* from genomic analysis of intestinal *E. coli* of Marmota himalayana in Qinghai-Tibet plateau of China. *Emerging Microbes and Infections*, *5*(e122), pp. 1–9. doi.org/10.1038/emi.2016.122

Lu, S., Zhang, X., Zhu, Y., Kim, K. S., Yang, J., & Jin, Q. (2011). Complete genome sequence of the neonatal-meningitis-associated *Escherichia coli* strain CE10. *Journal of Bacteriology*, *193*(24), pp. 7005. doi.org/10.1128/JB.06284-11

Luethy, P. M., Huynh, S., Ribardo, D. A., Winter, S. E., Parker, C. T., & Hendrixson, D. R. (2017). Microbiota-derived short-chain fatty acids modulate expression of *Campylobacter jejuni* determinants required for commensalism and virulence. *MBio*, *8*(3), pp. e00407-17. https://doi.org/10.1128/mBio

Maas, W. K., & Davis, B. D. (1950). Interference by D-serine and L-aspartic acid with pantothenate synthesis in *Escherichia coli*. *Pantothenate Studies*, 60(6), pp. 733–745. doi: 10.1128/jb.60.6.733-745.1950.

MacQuarrie, K. L., Fong, A. P., Morse, R. H., & Tapscott, S. J. (2011). Genome-wide transcription factor binding: Beyond direct target regulation. *Trends in Genetics, 27*(4), pp. 141–148. doi.org/10.1016/j.tig.2011.01.001.

Maddocks, S. E., & Oyston, P. C. F. (2008). Structure and function of the LysR-type transcriptional regulator (LTTR) family proteins. *Microbiology, 154*(12), pp. 3609–3623. doi.org/10.1099/mic.0.2008/022772-0

Magnusson, L. U., Farewell, A., & Nyström, T. (2005). ppGpp: A global regulator in *Escherichia coli*. *Trends in Microbiology*, *13*(5), pp. 236–242. doi.org/10.1016/j.tim.2005.03.008

Manges, A. R., Geum, H. M., Guo, A., Edens, T. J., Fibke, C. D., & Pitout, J. D. D. (2019). Global extraintestinal pathogenic *Escherichia coli* (Expec) lineages. *Clinical Microbiology Reviews*, *32*(3), pp. e00135-18. doi.org/10.1128/CMR.00135-18

Mann, K., & Jackson, M. A. (2008). Meningitis. *Pediatrics in Review*, *29*(12), pp. 417–430. doi.org/10.1542/pir.29-12-417

Mann, S., & Chen, Y. P. P. (2010). Bacterial genomic G + C composition-eliciting environmental adaptation. *Genomics*, *95*, pp. 7–15. https://doi.org/10.1016/j.ygeno.2009.09.002

Martínez-Antonio, A., & Collado-Vides, J. (2003). Identifying global regulators in transcriptional regulatory networks in bacteria. *Current Opinion in Microbiology*, *6*(5), pp. 482–489. doi.org/10.1016/j.mib.2003.09.002

Martinez, J. J., Mulvey, M. A., Schilling, J. D., Pinkner, J. S., & Hultgren, S. J. (2000). Type 1 pilus-mediated bacterial invasion of bladder epithelial cells. *The EMBO Journal*, *19*(12), pp. 2803–2812. doi: 10.1093/emboj/19.12.2803.

Maruvada, R., & Kim, K. S. (2012). IbeA and OmpA of *Escherichia coli* K1 exploit Rac1 activation for invasion of human brain microvascular endothelial cells. *Infection and Immunity*, *80*(6), pp. 2035–2041. doi.org/10.1128/IAI.06320-11

Marx, V. (2019). What to do about those immunoprecipitation blues. *Nature Methods*, *16*(4), pp. 289–292. doi.org/10.1038/s41592-019-0365-3

Mathew, R., & Chatterji, D. (2006). The evolving story of the omega subunit of bacterial RNA polymerase. *Trends in Microbiology*, *14*(10), pp. 450–455. doi.org/10.1016/j.tim.2006.08.002

Mazzulli, T. (2012). Diagnosis and management of simple and complicated urinary tract infections (UTIs). *The Canadian Journal of Urology*, *19*, pp. 42–48. PMID: 23089347.

McDaniel, T. K., Jarvis, K. G., Donnenberg, M. S., & Kaper, J. B. (1995). A genetic locus of enterocyte effacement conserved among diverse enterobacterial pathogens. *Proceedings of the National Academy of Sciences USA*, *92*(5), 1664–1668. doi: 10.1073/pnas.92.5.1664.

McFall, E. (1964). Genetic structure of the D-serine deaminase system of *Escherichia coli*. *Journal of Molecular Biology*, *9*(3), 746–753. doi.org/10.1016/S0022-2836(64)80179-0

McFall, E. (1967). Dominance studies with stable merodiploids in the D-Serine

deaminase system of *Escherichia coli* K-12. *Journal of Bacteriology*, *94*(6), pp. 1982–1988. doi: 10.1128/jb.94.6.1982-1988.1967.

McFall, E., & Heincz, M. C. (1983). Identification and control of synthesis of the *dsdC* activator protein. *Journal of Bacteriology*, *153*(2), pp. 872–877. doi: 10.1128/jb.153.2.872-877.1983.

McFall, S. M., Chugani, S. A., & Chakrabarty, A. M. (1998). Transcriptional activation of the catechol and chlorocatechol operons: variations on a theme. *Gene*, *223*, pp. 257–267. doi: 10.1016/s0378-1119(98)00366-7.

McFall, S. M., Klem, T. J., Fujita, N., Ishihama, A., & Chakrabarty, A. M. (1997). DNase I footprinting, DNA bending and *in vitro* transcription analyses of ClcR and CatR interactions with the *clcABD* promoter: evidence of a conserved transcriptional activation mechanism. *Molecular Microbiology*, *24*(5), pp. 965–976. doi: 10.1046/j.1365-2958.1997.4041763.x.

Mejía-Almonte, C., Busby, S. J. W., Wade, J. T., van Helden, J., Arkin, A. P., Stormo, G. D., Eilbeck, K., Palsson, B. O., Galagan, J. E., & Collado-Vides, J. (2020). Redefining fundamental concepts of transcription initiation in bacteria. *Nature Reviews Genetics*, *21*(11), pp. 699–714. doi.org/10.1038/s41576-020-0254-8

Melekos, M. D., & Naber, K. G. (2000). Complicated urinary tract infections. *International Journal of Antimicrobial Agents*, *15*(4), pp. 247-256. doi: 10.1016/s0924-8579(00)00168-0.

Mellies, J. L., Elliott, S. J., Sperandio, V., Donnenberg, M. S., & Kaper, J. B. (1999). The Per regulon of enteropathogenic *Escherichia coli* : identification of a regulatory cascade and a novel transcriptional activator, the locus of enterocyte effacement (LEE)-encoded regulator (Ler). *Molecular Microbiology*, *33*(2), pp. 296–306. doi: 10.1046/j.1365-2958.1999.01473.x.

Merrick, M. J. (1993). In a class of its own - the RNA polymerase sigma factor 54. *Molecular Microbiology*, *10*(5), pp. 903–909. doi: 10.1111/j.1365-2958.1993.tb00961.x.

Miajlovic, H., & Smith, S. G. (2014). Bacterial self-defence: How *Escherichia coli* evades serum killing. *FEMS Microbiology Letters*, *354*, pp. 1–9. doi.org/10.1111/1574-6968.12419.

Mikkelsen, T. S., Ku, M., Jaffe, D. B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T. K., Koche, R. P., Lee, W., Mendenhall, E., O'Donovan, A., Presser, A., Russ, C., Xie, X., Meissner, A., Wernig, M., Jaenisch, R., … Bernstein, B. E. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature*, *448*(7153), pp. 553–560. doi.org/10.1038/nature06008

Mitrophanov, A. Y., & Groisman, E. A. (2008). Signal integration in bacterial two-component regulatory systems. *Genes and Development*, *22*(19), pp. 2601–2611. doi.org/10.1101/gad.1700308

Mobley, H. L. T., Green, D. M., Trifillis, A. L., Johnson, D. E., Chippendale, G. R., Lockatell, C. V., Jones, B. D., & Warren, J. W. (1990). Pyelonephritogenic *Escherichia coli* and killing of cultured human renal proximal tubular epithelial cells: role of hemolysin in some strains. *Infection and Immunity*, *58*(5), pp. 1281–1289. doi: 10.1128/iai.58.5.1281-1289.1990.

Monroe, R. S., Ostrowski, J., Hryniewicz, M. M., & Kredich, N. M. (1990). *In vitro* interactions of CysB protein with the *cysK* and *cysJIH* promoter regions of *Salmonella typhimurium*. *Journal of Bacteriology*, *172*(12), pp. 6919–6929. doi: 10.1128/jb.172.12.6919-6929.1990.

Montso, P. K., Mlambo, V., & Ateba, C. N. (2019). Characterization of lytic bacteriophages infecting multidrug-resistant shiga toxigenic atypical *Escherichia coli* O177 strains isolated from cattle feces. *Frontiers in Public Health*, *7*(355), pp. 1-13. doi.org/10.3389/fpubh.2019.00355

Moon, H. W., Whipp, S. C., Argenzio, R. A., Levine, M. M., & Giannella, R. A. (1983). Attaching and effacing activities of rabbit and human Enteropathogenic *Escherichia coli* in pig and rabbit intestines. *Infection and Immunity*, *41*(3), pp. 1340–1351. doi: 10.1128/iai.41.3.1340-1351.1983.

Moon, K., Bonocora, R. P., Kim, D. D., Chen, Q., Wade, J. T., Stibitz, S., & Hinton, D. M. (2017). The BvgAS Regulon of *Bordetella pertussis*. *MBio*, *8*(5), pp. e01526-17. doi.org/10.1128/mBio

Mordhorst, I. L., Claus, H., Ewers, C., Lappann, M., Schoen, C., Elias, J., Batzilla, J., Dobrindt, U., Wieler, L. H., Bergfeld, A. K., Mühlenhoff, M., & Vogel, U. (2009). O-acetyltransferase gene *neuO* is segregated according to phylogenetic background

and contributes to environmental desiccation resistance in *Escherichia coli* K1. *Environmental Microbiology*, *11*(12), pp. 3154–3165. doi.org/10.1111/j.1462-2920.2009.02019.x

Moritz, R. L., & Welch, R. A. (2006). The *Escherichia coli argW-dsdCXA* genetic island is highly variable, and *E. coli* K1 strains commonly possess two copies of *dsdCXA*. *Journal of Clinical Microbiology*, *44*(11), pp. 4038–4048. doi.org/10.1128/JCM.01172-06

Mulvey, M. A., Lopez-Boado, Y. S., Wilson, C. L., Roth, R., Parks, W. C., Heuser, J., & Hultgren, S. J. (1998). Induction and evasion of host defences by Type 1-piliated Uropathogenic *Escherichia coli*. *Science*, *282*(5393), pp. 1494–1497. doi: 10.1126/science.282.5393.1494.

Murakami, K. S. (2015). Structural biology of bacterial RNA polymerase. *Biomolecules*, *5*(2), pp. 848–864. doi.org/10.3390/biom5020848.

Muraoka, S., Okumura, R., Ogawa, N., Nonaka, T., Miyashita, K., & Senda, T. (2003). Crystal Structure of a full-length LysR-type transcriptional regulator, CbnR: Unusual combination of two subunit forms and molecular bases for causing and changing DNA bend. *Journal of Molecular Biology*, *328*(3), pp. 555–566. doi.org/10.1016/S0022-2836(03)00312-7

Murray, G. L., Attridge, S. R., & Morona, R. (2006). Altering the length of the lipopolysaccharide O antigen has an impact on the interaction of *Salmonella enterica* serovar typhimurium with macrophages and complement. *Journal of Bacteriology*, *188*(7), pp. 2735–2739. doi.org/10.1128/JB.188.7.2735-2739.2006

Muto, A., & Osawa, S. (1987). The guanine and cytosine content of genomic DNA and bacterial evolution. *Proceedings of the National Academy of Sciences USA*, *84*(1), 166–169. doi: 10.1073/pnas.84.1.166.

Najafi, A., Hasanpour, M., Askary, A., Aziemzadeh, M., & Hashemi, N. (2018). Distribution of pathogenicity island markers and virulence factors in new phylogenetic groups of uropathogenic *Escherichia coli* isolates. *Folia Microbiologica*, *63*, pp. 335–343. doi.org/10.1007/s12223-017-0570-3

Navarro-Garcia, F., Ruiz-Perez, F., Cataldi, Á., & Larzábal, M. (2019). Type VI secretion system in pathogenic *Escherichia coli*: Structure, role in virulence, and acquisition.

Frontiers in Microbiology, *10*(1965), pp. 1-17. doi.org/10.3389/fmicb.2019.01965

Nishikawa, T. (2011). Analysis of free D-serine in mammals and its biological relevance. *Journal of Chromatography B*, *879*(29), pp. 3169–3183. doi.org/10.1016/j.jchromb.2011.08.030

Njoroge, J. W., Nguyen, Y., Curtis, M. M., Moreira, C. G., & Sperandio, V. (2012). Virulence meets metabolism: Cra and KdpE gene regulation in enterohemorrhagic *Escherichia coli*. *MBio*, *3*(5), pp. e00280-12. doi.org/10.1128/mBio.00280-12

Nørregaard-Madsen, M., McFall, E., & Valentin-Hansen, P. (1995). Organization and transcriptional regulation of the *Escherichia coli* K-12 D-Serine tolerance locus. *Journal of Bacteriology*, *177*(22), pp. 6456–6461. doi: 10.1128/jb.177.22.6456-6461.1995.

Nougayrède, J-P., Homburg, S., Taieb, F., Boury, M., Brzuszkiewicz, E., Gottschalk, G., Buchrieser, C., Hacker, J., Dobrindt, U., & Oswald, E. (2006). *Escherichia coli* induces DNA double-strand breaks in eukaryotic cells. *Science*, *313*(5788), pp. 848–851. doi.org/10.1126/science.1128439

O'Boyle, N., Turner, N. C. A., Roe, A. J., & Connolly, J. P. R. (2020). Plastic circuits: regulatory flexibility in fine tuning pathogen success. *Trends in Microbiology*, *28*(5), pp. 360–371. doi.org/10.1016/j.tim.2020.01.002

O'Meara, T. R., & Alspaugh, J. A. (2012). The *Cryptococcus neoformans* capsule: A sword and a shield. *Clinical Microbiology Reviews*, *25*(3), pp. 387–408. doi.org/10.1128/CMR.00001-12

Oliveros, J. (2007). *Venny. An interactive tool for comparing lists with Venn's diagrams*. Https://Bioinfogp.Cnb.Csic.Es/Tools/Venny/Index.Html.

Opal, S. M. (2014). Significance of sialic acid in *Klebsiella pneumoniae* K1 capsules. *Virulence*, *5*(6), pp. 648–649. doi.org/10.4161/viru.34349

Orskov, F., Orskov, I., Sutton, A., Schneerson, R., Lin, W., Egan, W., Hoff, E., & Robbins, J. B. (1979). Form variation in *Escherichia coli* K1: Determined by O-acetylation of the capsular polysaccheride. *Journal of Experimental Medicine*, *149*(3), pp. 669–685. doi: 10.1084/jem.149.3.669.

Pacheco, A. R., Munera, D., Waldor, M. K., Sperandio, V., & Ritchie, J. M. (2012). Fucose sensing regulates bacterial intestinal colonization. *Nature*, *492*, pp. 113–117. doi.org/10.1038/nature11623

Paget, M. S. (2015). Bacterial sigma factors and anti-sigma factors: Structure, function and distribution. *Biomolecules*, *5*(3), pp. 1245–1265. doi.org/10.3390/biom5031245

Pareja, E., Pareja-Tobes, P., Manrique, M., Pareja-Tobes, E., Bonal, J., & Tobes, R. (2006). ExtraTrain: A database of extragenic regions and transcriptional information in prokaryotic organisms. *BMC Microbiology*, *6*(29), pp. 1–10. doi.org/10.1186/1471-2180-6-29

Pérez-Rueda, E., & Collado-Vides, J. (2000). The repertoire of DNA-binding transcriptional regulators in *Escherichia coli* K-12. *Nucleic Acids Research*, *28*(8), pp. 1838–1847. doi: 10.1093/nar/28.8.1838.

Perez, J. C., & Groisman, E. A. (2009a). Transcription factor function and promoter architecture govern the evolution of bacterial regulons. *Proceedings of the National Academy of Sciences USA*, *106*(11), pp. 4319–4324. doi: 10.1073/pnas.0810343106.

Perez, J. C., & Groisman, E. A. (2009b). Evolution of Transcriptional Regulatory Circuits in Bacteria. *Cell*, *138*(2), pp. 233–244. doi.org/10.1016/j.cell.2009.07.002

Pickrell, J. K., Gaffney, D. J., Gilad, Y., & Pritchard, J. K. (2011). False positive peaks in ChIP-seq and other sequencing-based functional assays caused by unannotated high copy number regions. *Bioinformatics*, *27*(15), pp. 2144–2146. doi.org/10.1093/bioinformatics/btr354

Pires, D. P., Oliveira, H., Melo, L. D. R., Sillankorva, S., & Azeredo, J. (2016). Bacteriophage-encoded depolymerases: their diversity and biotechnological applications. *Applied Microbiology and Biotechnology*, *100*(5), pp. 2141–2151. doi.org/10.1007/s00253-015-7247-0

Polin, R. A., & Harris, M. C. (2001). Neonatal bacterial meningitis. *Seminars in Neonatology*, *6*(2), pp. 157–172. doi.org/10.1053/siny.2001.0045

Pollegioni, L., & Sacchi, S. (2010). Metabolism of the neuromodulator D-serine. *Cellular and Molecular Life Sciences*, *67*(14), pp. 2387–2404. doi.org/10.1007/s00018-010-

0307-9

Poncet, S., Milohanic, E., Mazé, A., Abdallah, J. N., Aké, F., Larribe, M., Deghmane, A.-E., Taha, M-K., Dozot, M., De Bolle, X., Letesson, J. J., & Deutscher, J. (2009). Correlations between carbon metabolism and virulence in bacteria. In M. Collin & R. Schuch (Eds.), *Bacterial sensing and signalling*, *16*, pp. 88–102. doi: 10.1159/000219374.

Poulsen, L. K., Lan, F., Kristensen, C. S., Hobolth, P., Molin, S., & Krogfelt, K. A. (1994). Spatial distribution of *Escherichia coli* in the mouse large intestine inferred from rRNA *in situ* hybridization. *Infection and Immunity*, *62*(11), pp. 5191–5194. doi: 10.1128/iai.62.11.5191-5194.1994.

Prasadarao, N. V. (2002). Identification of *Escherichia coli* outer membrane protein A receptor on human brain microvascular endothelial cells. *Infection and Immunity*, *70*(8), pp. 4556–4563. doi.org/10.1128/IAI.70.8.4556-4563.2002

Prasadarao, N. V., Wass, C. A., Weiser, J. N., Stins, M. F., Huang, S-H., & Kim, K. S. (1996). Outer membrane protein A of *Escherichia coli* contributes to invasion of brain microvascular endothelial cells. *Infection and Immunity*, *64*(1), pp. 146–153. doi: 10.1128/iai.64.1.146-153.1996.

Prokhorov, N. S., Riccio, C., Zdorovenko, E. L., Shneider, M. M., Browning, C., Knirel, Y. A., Leiman, P. G., & Letarov, A. V. (2017). Function of bacteriophage G7C esterase tailspike in host cell adsorption. *Molecular Microbiology*, *105*(3), pp. 385–398. doi.org/10.1111/mmi.13710

Raetz, C. R. H., & Whitfield, C. (2002). Lipopolysaccharide endotoxins. *Annual Review of Biochemistry*, *71*, pp. 635–700. doi.org/10.1146/annurev.biochem.71.110601.135414

Rasko, D. A., Rosovitz, M. J., Myers, G. S. A., Mongodin, E. F., Fricke, W. F., Gajer, P., Crabtree, J., Sebaihia, M., Thomson, N. R., Chaudhuri, R., Henderson, I. R., Sperandio, V., & Ravel, J. (2008). The pangenome structure of *Escherichia coli*: Comparative genomic analysis of *E. coli* commensal and pathogenic isolates. *Journal of Bacteriology*, *190*(20), pp. 6881–6893. doi.org/10.1128/JB.00619-08.

Rasko, D. A., & Sperandio, V. (2010). Anti-virulence strategies to combat bacteria-mediated disease. *Nature Reviews Drug Discovery*, *9*(2), pp. 117–128.

doi.org/10.1038/nrd3013.

Ray, B., Ballal, A., & Manna, A. C. (2009). Transcriptional variation of regulatory and virulence genes due to different media in *Staphylococcus aureus*. *Microbial Pathogenesis*, *47*(2), pp. 94–100. doi.org/10.1016/j.micpath.2009.05.001.

Reigstad, C. S., Hultgren, S. J., & Gordon, J. I. (2007). Functional genomic studies of uropathogenic *Escherichia coli* and host urothelial cells when intracellular bacterial communities are assembled. *Journal of Biological Chemistry*, *282*(29), pp. 21259–21267. doi.org/10.1074/jbc.M611502200

Reister, M., Hoffmeier, K., Krezdorn, N., Rotter, B., Liang, C., Rund, S., Dandekar, T., Sonnenborn, U., & Oelschlaeger, T. A. (2014). Complete genome sequence of the Gram-negative probiotic *Escherichia coli* strain Nissle 1917. *Journal of Biotechnology*, *187*, pp. 106–107. doi.org/10.1016/j.jbiotec.2014.07.442

Ren, C. P., Chaudhuri, R. R., Fivian, A., Bailey, C. M., Antonio, M., Barnes, W. M., & Pallen, M. J. (2004). The ETT2 gene cluster, encoding a second type III secretion system from *Escherichia coli*, is present in the majority of strains but has undergone widespread mutational attrition. *Journal of Bacteriology*, *186*(11), pp. 3547–3560. doi.org/10.1128/JB.186.11.3547-3560.2004

Robbins, J. B., McCracken Jr, G. H., Gotschlich, E. C., Orskov, F., Orskov, I., & Hanson, L. A. (1974). *Escherichia coli* K1 capsular polysaccharide associated with neonatal meningitis. *The New England Journal of Medicine*, *290*(22), pp. 1216–1220. doi: 10.1056/NEJM197405302902202.

Roberts, J. W. (2019). Mechanisms of bacterial transcription termination. *Journal of Molecular Biology*, *431*(20), 4030–4039. doi.org/10.1016/j.jmb.2019.04.003.

Robertson, G., Hirst, M., Bainbridge, M., Bilenky, M., Zhao, Y., Zeng, T., Euskirchen, G., Bernier, B., Varhol, R., Delaney, A., Thiessen, N., Griffith, O. L., He, A., Marra, M., Snyder, M., & Jones, S. (2007). Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nature Methods*, *4*(8), pp. 651–657. doi.org/10.1038/nmeth1068.

Roesch, P. L., Redford, P., Batchelet, S., Moritz, R. L., Pellett, S., Haugen, B. J., Blattner, F. R., & Welch, R. A. (2003). Uropathogenic *Escherichia coli* use D-serine deaminase to modulate infection of the murine urinary tract. *Molecular*

Microbiology, *49*(1), pp. 55–67. doi.org/10.1046/j.1365-2958.2003.03543.x

Roos, V., & Klemm, P. (2006). Global gene expression profiling of the asymptomatic bacteriuria *Escherichia coli* strain 83972 in the human urinary tract. *Infection and Immunity*, *74*(6), pp. 3565–3575. doi.org/10.1128/IAI.01959-05

Russo, T. A., & Johnson, J. R. (2000). Proposal for a new inclusive designation for extraintestinal pathogenic isolates of *Escherichia coli*: ExPEC. *The Journal of Infectious Diseases*, *181*(5), 1753–1754. doi: 10.1086/315418.

Russo, T. A., & Johnson, J. R. (2003). Medical and economic impact of extraintestinal infections due to *Escherichia coli:* focus on an increasingly important endemic problem. *Microbes and Infection*, *5*, pp. 449–456. doi: 10.1016/s1286-4579(03)00049-2.

Sachdev, D., & Chirgwin, I. M. (2000). Fusions to maltose-binding protein: control of folding and solubility in protein purification. *Methods in Enzymology*, *326*, pp. 312–321. doi: 10.1016/s0076-6879(00)26062-x.

Sanchez-Vazquez, P., Dewey, C. N., Kitten, N., Ross, W., & Gourse, R. L. (2019). Genome-wide effects on *Escherichia coli* transcription from ppGpp binding to its two sites on RNA polymerase. *Proceedings of the National Academy of Sciences of the USA*, *116*(17), pp. 8310–8319. doi.org/10.1073/pnas.1819682116.

Santangelo, T. J., & Roberts, J. W. (2004). Forward translocation is the natural pathway of RNA release at an intrinsic terminator. *Molecular Cell*, *14*(1), 117–126. doi: 10.1016/s1097-2765(04)00154-6.

Sarkar, S., Ulett, G. C., Totsika, M., Phan, M. D., & Schembri, M. A. (2014). Role of capsule and O antigen in the virulence of uropathogenic *Escherichia coli*. *PLoS ONE*, *9*(4), pp. e94786. doi.org/10.1371/journal.pone.0094786

Sasabe, J., Miyoshi, Y., Rakoff-Nahoum, S., Zhang, T., Mita, M., Davis, B. M., Hamase, K., & Waldor, M. K. (2016). Interplay between microbial D-amino acids and host D-amino acid oxidase modifies murine mucosal defence and gut microbiota. *Nature Microbiology*, *1*, pp. 16125. doi.org/10.1038/nmicrobiol.2016.125.

Schell, M. A. (1993). Molecular biology of the LysR family of transcriptional regulators. *Annual Review of Microbiology*, *47*, pp. 597–626. doi:

10.1146/annurev.mi.47.100193.003121.

Scholl, D., Adhya, S., & Merril, C. (2005). *Escherichia coli* K1's capsule is a barrier to bacteriophage T7. *Applied and Environmental Microbiology*, *71*(8), pp. 4872–4874. doi.org/10.1128/AEM.71.8.4872-4874.2005

Schrader, S. M., Vaubourgeix, J., & Nathan, C. (2020). Biology of antimicrobial resistance and approaches to combat it. *Sci. Transl. Med*, *12*, pp. eaaz6992. doi: 10.1126/scitranslmed.aaz6992.

Schwechheimer, C., & Kuehn, M. J. (2015). Outer-membrane vesicles from Gram-negative bacteria: Biogenesis and functions. *Nature Reviews Microbiology, 13*(10), pp. 605–619. doi.org/10.1038/nrmicro3525.

Secher, T., Samba-Louaka, A., Oswald, E., & Nougayrède, J. P. (2013). *Escherichia coli* producing colibactin triggers premature and transmissible senescence in mammalian cells. *PLoS ONE*, *8*(10), pp. e77157. doi.org/10.1371/journal.pone.0077157

Seo, S. W., Kim, D., Szubin, R., & Palsson, B. O. (2015). Genome-wide reconstruction of OxyR and SoxRS transcriptional regulatory networks under oxidative stress in *Escherichia coli* K-12 MG1655. *Cell Reports*, *12*, pp. 1289–1299. doi.org/10.1016/j.celrep.2015.07.043

Shah, C., Baral, R., Bartaula, B., & Shrestha, L. B. (2019). Virulence factors of uropathogenic *Escherichia coli* (UPEC) and correlation with antimicrobial resistance. *BMC Microbiology*, *19*(204), pp. 1–6. doi.org/10.1186/s12866-019-1587-3

Shimada, T., Ishihama, A., Busby, S. J. W., & Grainger, D. C. (2008). The *Escherichia coli* RutR transcription factor binds at targets within genes as well as intergenic regions. *Nucleic Acids Research*, *36*(12), pp. 3950–3955. doi.org/10.1093/nar/gkn339

Shimizu, K. (2013). Regulation systems of bacteria such as *Escherichia coli* in response to nutrient limitation and environmental stresses. *Metabolites*, *4*, pp. 1–35. doi.org/10.3390/metabo4010001

Silver, R., & Vimr, E. R. (1990). Polysialic acid capsule of *Escherichia coli* K1. In B.

Iglewski & V. Clark (Eds.), *Molecular Basis of Bacterial Pathogenesis, 11,* pp. 39–54. Academic Press Ltd, London, NW1 7DX.

Singh, S. S., Singh, N., Bonocora, R. P., Fitzgerald, D. M., Wade, J. T., & Grainger, D. C. (2014). Widespread suppression of intragenic transcription initiation by H-NS. *Genes and Development*, *28*, pp. 214–219. doi.org/10.1101/gad.234336.113

Slater, S. L., Sågfors, A. M., Pollard, D. J., Ruano-Gallego, D., & Frankel, G. (2018). The type III secretion system of pathogenic *Escherichia coli*. *Current Topics in Microbiology and Immunology. Escherichia coli: A versatile pathogen, 416*, pp. 51–72. doi.org/10.1007/82_2018_116

Stark, R., Grzelak, M., & Hadfield, J. (2019). RNA sequencing: the teenage years. *Nature Reviews Genetics, 20*(11), pp. 631–656). doi.org/10.1038/s41576-019-0150-2

Starmer, J., & Magnuson, T. (2016). Detecting broad domains and narrow peaks in ChIP-seq data with hiddenDomains. *BMC Bioinformatics*, *17*(144), pp. 1-10. doi.org/10.1186/s12859-016-0991-z

Starosta, A. L., Lassak, J., Jung, K., & Wilson, D. N. (2014). The bacterial translation stress response. *FEMS Microbiology Reviews*, *38*(6), pp. 1172–1201. doi.org/10.1111/1574-6976.12083

Stephens, D. S., Spellman, P. A., & Swartley, J. S. (1993). Effect of the (a2-8)-linked polysialic acid capsule on adherence of *Neisseria meningitidis* to human mucosal cells. *Journal of Infectious Diseases*, *167*, pp. 475–478. doi: 10.1093/infdis/167.2.475.

Stone, N. L., England, T. J., & O'Sullivan, S. E. (2019). A novel transwell blood brain barrier model using primary human cells. *Frontiers in Cellular Neuroscience*, *13*(230), pp. 1-11. doi.org/10.3389/fncel.2019.00230

Stummeyer, K., Schwarzer, D., Claus, H., Vogel, U., Gerardy-Schahn, R., & Mühlenhoff, M. (2006). Evolution of bacteriophages infecting encapsulated bacteria: Lessons from *Escherichia coli* K1-specific phages. *Molecular Microbiology*, *60*(5), pp. 1123–1135. doi.org/10.1111/j.1365-2958.2006.05173.x

Sullivan, M. J., Petty, N. K., & Beatson, S. A. (2011). Easyfig: A genome comparison

visualizer. *Bioinformatics*, *27*(7), pp. 1009–1010. doi.org/10.1093/bioinformatics/btr039

Suzuki, M., Sujino, T., Chiba, S., Harada, Y., Goto, M., Takahashi, R., Mita, M., Hamase, K., Kanai, T., Ito, M., Waldor, M. K., Yasui, M., & Sasabe, J. (2021). Host-microbe cross-talk governs amino acid chirality to regulate survival and differentiation of B cells. *Science Advances*, *7*(10), pp. eabd6480. doi: 10.1126/sciadv.abd6480.

Teichmann, S. A., & Babu, M. M. (2004). Gene regulatory network growth by duplication. *Nature Genetics*, *36*(5), pp. 492–496. doi.org/10.1038/ng1340

Teichmann, S. A., Park, J., & Chothia, C. (1998). Structural assignments to the *Mycoplasma genitalium* proteins show extensive gene duplications and domain rearrangements. *Proceedings of the National Academy of Sciences USA*, *95*, pp. 14658–14663. doi: 10.1073/pnas.95.25.14658.

Tenaillon, O., Skurnik, D., Picard, B., & Denamur, E. (2010). The population genetics of commensal *Escherichia coli*. *Nature Reviews Microbiology*, *8*, pp. 207–217. doi.org/10.1038/nrmicro2298

Terlizzi, M. E., Gribaudo, G., & Maffei, M. E. (2017). UroPathogenic *Escherichia coli* (UPEC) infections: Virulence factors, bladder responses, antibiotic, and non-antibiotic antimicrobial strategies. *Frontiers in Microbiology*, *8*(1566), pp. 1-23. doi.org/10.3389/fmicb.2017.01566

Thomas-White, K., Forster, S. C., Kumar, N., Van Kuiken, M., Putonti, C., Stares, M. D., Hilt, E. E., Price, T. K., Wolfe, A. J., & Lawley, T. D. (2018). Culturing of female bladder bacteria reveals an interconnected urogenital microbiota. *Nature Communications*, *9*(1557), pp. 1-7. doi.org/10.1038/s41467-018-03968-5

Touchon, M., Bernheim, A., & Rocha, E. P. C. (2016). Genetic and life-history traits associated with the distribution of prophages in bacteria. *ISME Journal*, *10*(11), pp. 2744–2754. doi.org/10.1038/ismej.2016.47

Touchon, M., Hoede, C., Tenaillon, O., Barbe, V., Baeriswyl, S., Bidet, P., Bingen, E., Bonacorsi, S., Bouchier, C., Bouvet, O., Calteau, A., Chiapello, H., Clermont, O., Cruveiller, S., Danchin, A., Diard, M., Dossat, C., El Karoui, M., Frapy, E., … Denamur, E. (2009). Organised genome dynamics in the *Escherichia coli* species

results in highly diverse adaptive paths. *PLoS Genetics*, *5*(1), pp. e1000344. doi.org/10.1371/journal.pgen.1000344

Tsai, C-M., & Frasch, C. E. (1982). A sensitive silver stain for detecting lipopolysaccharides in polyacrylamide gels. *Analytical Biochemistry*, *119*, pp. 115–119. doi: 10.1016/0003-2697(82)90673-x.

Turner, N. C. A., Connolly, J. P. R., & Roe, A. J. (2018). Control freaks—signals and cues governing the regulation of virulence in attaching and effacing pathogens. *Biochemical Society Transactions*, *47*(1), pp. 229–238. doi.org/10.1042/BST20180546

Tzeng, Y. L., Thomas, J., & Stephens, D. S. (2016). Regulation of capsule in *Neisseria meningitidis*. *Critical Reviews in Microbiology*, *42*(5), pp. 759–772. doi.org/10.3109/1040841X.2015.1022507

Uzureau, S., Lemaire, J., Delaive, E., Dieu, M., Gaigneaux, A., Raes, M., De Bolle, X., & Letesson, J. J. (2010). Global analysis of quorum sensing targets in the intracellular pathogen *Brucella melitensis* 16 M. *Journal of Proteome Research*, *9*(6), pp. 3200–3217. doi.org/10.1021/pr100068p

Van Den Akker, W. M. R. (1998). Lipopolysaccharide expression within the genus *Bordetella*: influence of temperature and phase variation. *Microbiology*, *144*, pp. 1527–1535. doi: 10.1099/00221287-144-6-1527

Van Der Woude, M. W., & Bäumler, A. J. (2004). Phase and antigenic variation in bacteria. *Clinical Microbiology Reviews*, *17*(3), pp. 581–611. doi.org/10.1128/CMR.17.3.581-611.2004

Vannini, A., Pinatel, E., Costantini, P. E., Pelliciari, S., Roncarati, D., Puccio, S., De Bellis, G., Peano, C., & Danielli, A. (2017). Comprehensive mapping of the *Helicobacter pylori* NikR regulon provides new insights in bacterial nickel responses. *Scientific Reports*, *7*(45458), pp. 1–14. doi.org/10.1038/srep45458

Vejborg, R. M., Friis, C., Hancock, V., Schembri, M. A., & Klemm, P. (2010). A virulent parent with probiotic progeny: Comparative genomics of *Escherichia coli* strains CFT073, Nissle 1917 and ABU 83972. *Molecular Genetics and Genomics*, *283*(5), pp. 469–484. doi.org/10.1007/s00438-010-0532-9

Vermeulen, C., Cross, A., Byrne, W. R., & Zollinger2, A. W. (1988). Quantitative relationship between capsular content and killing of K1-encapsulated *Escherichia coli*. *Infection and Immunity*, *56*(10), pp. 2723–2730. doi: 10.1128/iai.56.10.2723-2730.1988.

Vihta, K. D., Stoesser, N., Llewelyn, M. J., Quan, T. P., Davies, T., Fawcett, N. J., Dunn, L., Jeffery, K., Butler, C. C., Hayward, G., Andersson, M., Morgan, M., Oakley, S., Mason, A., Hopkins, S., Wyllie, D. H., Crook, D. W., Wilcox, M. H., Johnson, A. P., … Walker, A. S. (2018). Trends over time in *Escherichia coli* bloodstream infections, urinary tract infections, and antibiotic susceptibilities in Oxfordshire, UK, 1998–2016: a study of electronic health records. *The Lancet Infectious Diseases*, *18*(10), pp. 1138–1149. doi.org/10.1016/S1473-3099(18)30353-0

Vimr, E., Steenbergen, S., & Cieslewicz, M. (1995). Biosynthesis of the polysialic acid capsule in *Escherichia coli* K1. *Journal of Industrial Microbiology*, *15*, pp. 352–360. doi: 10.1007/BF01569991.

Vörös, Z., Yan, Y., Kovari, D. T., Finzi, L., & Dunlap, D. (2017). Proteins mediating DNA loops effectively block transcription. *Protein Science*, *26*, pp. 1427–1438. doi.org/10.1002/pro.3156

Wade, J. T. (2015). Mapping transcription regulatory networks with ChIP-seq and RNA-seq. *Advances in Experimental Medicine and Biology*, *883*, pp. 119–134). doi.org/10.1007/978-3-319-23603-2_7

Wallenstein, A., Rehm, N., Brinkmann, M., Selle, M., Bossuet-Greif, N., Sauer, D., Bunk, B., Spröer, C., Wami, H. T., Homburg, S., Von Bünau, R., König, S., Nougayrède, J-P., Overmann, J., Oswald, E., Müller, R., & Dobrindt, U. (2020). ClbR Is the key transcriptional activator of colibactin gene expression in *Escherichia coli*. *MSphere*, *5*(4), pp. e00591-20. doi.org/10.1128/mSphere

Wang, I-N., Dykhuizen, D. E., & Slobodkin, L. B. (1996). The evolution of phage lysis timing. *Evolutionary Ecology*, *10*(5), pp. 545–558. doi:10.1007/BF01237884

Wang, X., & Quinn, P. J. (2010). Endotoxins: Lipopolysaccharides of Gram-negative bacteria. *Endotoxins: Structure, function and recognition, Subcellular Biochemistry, 53*, pp. 3–27. doi: 10.1007/978-90-481-9078-2_1.

Washburn, R. S., & Gottesman, M. E. (2015). Regulation of transcription elongation and termination. *Biomolecules*, *5*(2), pp. 1063–1078. doi.org/10.3390/biom5021063

Watts, R. E., Totsika, M., Challinor, V. L., Mabbett, A. N., Ulett, G. C., Voss, J. J. D., & Schembri, M. A. (2012). Contribution of siderophore systems to growth and urinary tract colonization of asymptomatic bacteriuria *Escherichia coli*. *Infection and Immunity*, *80*(1), pp. 333–344. doi.org/10.1128/IAI.05594-11

Waugh, D. S. (2011). An overview of enzymatic reagents for the removal of affinity tags. *Protein Expression and Purification*, *80*(2), pp. 283–293. doi.org/10.1016/j.pep.2011.08.005

Wei, Q., Le Minh, P. N., Dötsch, A., Hildebrand, F., Panmanee, W., Elfarash, A., Schulz, S., Plaisance, S., Charlier, D., Hassett, D., Häussler, S., & Cornelis, P. (2012). Global regulation of gene expression by OxyR in an important human opportunistic pathogen. *Nucleic Acids Research*, *40*(10), pp. 4320–4333. doi.org/10.1093/nar/gks017

Welch, R A, Burland, V., Plunkett III, G., Redford, P., Roesch, P., Rasko, D., Buckles, E. L., Liou, S.-R., Boutin, A., Hackett, J., Stroud, D., Mayhew, G. F., Rose, D. J., Zhou, S., Schwartz, D. C., Perna, N. T., Mobley, H. L. T., & Blattner, F. R. (2002). Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proceedings of the National Academy of Sciences USA*, *99*(26), pp. 17020-17024. doi: 10.1073/pnas.252529799.

Welch, R A. (2016). Uropathogenic *Escherichia coli*-associated exotoxins. *Microbiology Spectrum*, *4*(3), pp. 1–16. doi.org/10.1128/microbiolspec.uti-0011-2012

Whitfield, C., Heinrichs, D. E., Yethon, J. A., Amor, K. L., Monteiro, M. A., & Perry, M. B. (1999). Assembly of the R1-type core oligosaccharide of *Escherichia coli* lipopolysaccharide. *Journal of Endotoxin Research*, *5*(3), pp. 151-156. doi: 10.1074/jbc.273.45.29497.

Whitfield, C., & Roberts, I. S. (1999). Structure, assembly and regulation of expression of capsules in *Escherichia coli*. *Molecular Microbiology*, *31*(5), pp. 1307–1319. doi: 10.1046/j.1365-2958.1999.01276.x.

Wijetunge, D. S. S., Gongati, S., Debroy, C., Kim, K. S., Couraud, P. O., Romero, I. A., Weksler, B., & Kariyawasam, S. (2015). Characterizing the pathotype of neonatal

meningitis causing *Escherichia coli* (NMEC). *BMC Microbiology*, *15*(211), pp. 1–15. doi.org/10.1186/s12866-015-0547-9

Wilson, K. S., & Von Hippel, P. H. (1995). Transcription termination at intrinsic terminators: The role of the RNA hairpin. *Proceedings of the National Academy of Sciences USA*, *92*, pp. 8793–8797. doi: 10.1073/pnas.92.19.8793.

Wolf, S. G., Frenkiel, D., Arad, T., Finkel, S. E., Kolter, R., & Minsky, A. (1999). DNA protection by stress-induced biocrystallization. *Nature*, *400*, pp. 83–85. doi: 10.1038/21918.

Xie, Y., Kim, K. J., & Kim, K. S. (2004). Current concepts on *Escherichia coli* K1 translocation of the blood-brain barrier. *FEMS Immunology and Medical Microbiology*, *42*(3), pp. 271–279. doi.org/10.1016/j.femsim.2004.09.001

Yao, Y., Xie, Y., & Kim, K. S. (2006). Genomic comparison of *Escherichia coli* K1 strains isolated from the cerebrospinal fluid of patients with meningitis. *Infection and Immunity*, *74*(4), pp. 2196–2206. doi.org/10.1128/IAI.74.4.2196-2206.2006

Yao, Y., Xie, Y., Perace, D., Zhong, Y., Lu, J., Tao, J., Guo, X., & Kim, K. S. (2009). The type III secretion system is involved in the invasion and intracellular survival of *Escherichia coli* K1 in human brain microvascular endothelial cells. *FEMS Microbiology Letters*, *300*(1), pp. 18–24. doi.org/10.1111/j.1574-6968.2009.01763.x

Yarnell, W. S., & Roberts, J. W. (1999). Mechanism of intrinsic transcription termination and antitermination. *Science, 284*(5414), pp. 611–615. doi: 10.1126/science.284.5414.611.

Yethon, J. A., Vinogradov, E., Perry, M. B., & Whitfield, C. (2000). Mutation of the lipopolysaccharide core glycosyltransferase encoded by *waaG* destabilizes the outer membrane of *Escherichia coli* by interfering with core phosphorylation. *Journal of Bacteriology*, *182*(19), pp. 5620–5623. doi: 10.1128/JB.182.19.5620-5623.2000.

Young, C. L., Britton, Z. T., & Robinson, A. S. (2012). Recombinant protein expression and purification: A comprehensive review of affinity tags and microbial applications. *Biotechnology Journal*, *7*(5), pp. 620–634. doi.org/10.1002/biot.201100155

Zagon, J., Dehne, L-I., & Bogl, K.-W. (1994). D-amino acids in organisms and food. *Nutrition Research*, *14*(3), pp. 445–463. doi.org/10.1016/S0271-5317(05)80182-4

Zhang, C., An, H., Hu, J., Li, J., Zhang, W., Lan, X., Deng, H., & Zhang, J. (2021). MetR is a molecular adaptor for pneumococcal carriage in the healthy upper airway. *Molecular Microbiology, 116*(2), pp. 438-458. doi.org/10.1111/mmi.14724

Zhang, J. (2003). Evolution by gene duplication: An update. *Trends in Ecology and Evolution, 18*(6), pp. 292–298. doi.org/10.1016/S0169-5347(03)00033-8

Zhou, Y., Zhang, X., & Ebright, R. H. (1993). Identification of the activating region of catabolite gene activator protein (CAP): Isolation and characterization of mutants of CAP specifically defective in transcription activation. *Proceedings of National Academy of Sciences USA*, *90*, pp. 6081–6085. doi: 10.1073/pnas.90.13.6081.

Supplementary data

Table S-1 CFT073 binding sites (no D-serine).

			Replicate 1				Replicate 2			
Gene	Dist. Gene	Context	Co-ordinates	Centre	Length	*p*-value	Co-ordinates	Centre	Length	*p*-value
c0346	-54	Intergenic	322255..322761	322527	507	3.93E-15	322294..322729	322542	436	3.31E-03
c1208	-72	Intergenic	1164775..1165252	1165010	478	5.09E-14	1164795..1165240	1165052	446	1.83E-03
c3694	0	Intragenic	3525736..3526315	3526079	580	8.71E-10	3525893..3526313	3526125	421	9.48E-04
rfaJ	0	Intergenic	4234841..4235368	4235076	528	2.33E-09	4234927..4235349	4235116	423	3.55E-05
wzy	128	Intragenic	2399200..2399768	2399437	569	2.08E-08	2399020..2399493	2399305	474	5.34E-04
waaL	705	Intragenic	4230281..4230846	4230516	566	2.37E-08	4230274..4230719	4230531	446	2.39E-04
rfaY	0	Intragenic	4233801..4234394	4234038	594	2.38E-08	4233850..4234354	4234038	505	6.11E-03
c0327	0	Intergenic	299056..299527	299293	472	2.55E-08	299096..299478	299284	383	1.00E-02
yrhA	0	Intragenic	4022294..4022777	4022529	484	3.79E-08	4022325..4022703	4022514	379	3.13E-03
waaV	0	Intergenic	4232118..4232633	4232400	516	4.08E-08	4232132..4232518	4232334	387	5.49E-05
c3766	0	Intragenic	3600396..3600931	3600633	536	5.12E-08	3600475..3600858	3600670	384	7.93E-03
yhiM	0	Intergenic	4068883..4069381	4069147	499	3.79E-07	4068903..4069318	4069131	416	4.87E-04
c5209	-224	Intragenic	4964289..4964816	4964524	528	3.81E-07	4964007..4964569	4964196	563	3.35E-03
c4303	0	Intergenic	4082826..4083296	4083061	471	4.61E-07	4082833..4083261	4083022	429	1.00E-02
c5426	146	Intragenic	5171549..5172100	5171786	552	6.92E-07	5171634..5172097	5171909	464	1.00E-02
yhhZ	0	Intragenic	4021035..4021608	4021272	574	1.14E-06	4021051..4021435	4021247	385	4.68E-04
c0430	-159	Intergenic	415182..415654	415420	473	1.21E-06	415260..415780	415593	521	2.18E-03
c2897	0	Intergenic	2757208..2757733	2757445	526	1.24E-06	2757362..2757847	2757551	486	3.46E-04
c3302	0	Intergenic	3148558..3149079	3148843	522	2.08E-06	3148463..3148959	3148652	497	2.00E-02
c4739	0	Intergenic	4507151..4507687	4507453	537	2.12E-06	4507120..4507601	4507413	482	1.72E-03
ycdT	0	Intergenic	1125620..1126089	1125855	470	1.06E-05	1125678..1126058	1125871	381	3.89E-03
cydA	0	Intragenic	792719..793303	792954	585	1.58E-05	792750..793161	792939	412	2.00E-02

tauA	0	Intergenic	459923..460449	460158	527	1.90E-05	460046..460429	460241	384	4.47E-03
c4423	0	Intergenic	4204869..4205392	4205106	524	2.18E-05	4204877..4205275	4205087	399	3.05E-03
yicL	-71	Intragenic	4342419..4342916	4342653	498	2.20E-05	4342634..4343028	4342840	395	1.00E-02
yfcV	-19	Intergenic	2737741..2738222	2737988	482	3.19E-05	2737711..2738226	2738038	516	2.45E-03
arcA	0	Intragenic	5094971..5095445	5095211	475	3.54E-05	5094933..5095340	5095097	408	3.00E-02
c0080	0	Intragenic	74758..75274	75040	517	4.19E-05	74744..75153	74933	410	1.13E-03
c5435	0	Intergenic	5181592..5182228	5181992	637	4.72E-05	5181803..5182216	5182036	414	3.00E-02
tssB	0	Intergenic	3222515..3223024	3222750	510	4.94E-05	3222504..3223023	3222693	520	2.83E-03
asst	0	Intergenic	3618369..3618857	3618623	489	5.68E-05	3618438..3618843	3618655	406	1.80E-04
c4498	585	Intergenic	4282736..4283276	4282971	541	1.16E-04	4282822..4283234	4283011	413	6.68E-03
caiT	0	Intergenic	44606..45107	44841	502	1.33E-04	44597..45059	44871	463	9.97E-03
c0322	0	Intergenic	292356..292924	292591	569	1.94E-04	292333..292766	292522	434	5.12E-03
hemB	-41	Intergenic	465027..465546	465262	520	2.37E-04	464974..465388	465163	415	7.11E-03
c4214	0	Intergenic	3999566..4000053	3999801	488	2.67E-04	3999704..4000094	3999906	391	3.00E-03
cmtB	0	Intergenic	3372727..3373422	3373186	696	2.89E-04	3373068..3373500	3373314	433	2.58E-03
c0321	20	Intragenic	291198..291685	291451	488	3.28E-04	291150..291637	291449	488	8.83E-03
leuO	-115	Intergenic	87602..88079	87839	478	3.36E-04	87710..88087	87899	378	7.60E-03
yjjQ	-47	Intergenic	5191907..5192418	5192184	512	4.58E-04	5192005..5192417	5192191	413	2.00E-02
c0077	0	Intergenic	73819..74471	74056	653	4.59E-04	73931..74480	74292	550	8.17E-03
c3686	-151	Intergenic	3511416..3511940	3511704	525	4.66E-04	3511413..3511823	3511602	411	2.00E-02
c2383	0	Intergenic	2185354..2185881	2185645	528	5.22E-04	2185385..2185931	2185743	547	2.00E-02
c2405	95	Intragenic	2208165..2208658	2208424	494	5.42E-04	2208150..2208572	2208384	423	8.25E-03
hlyC	-412	Intergenic	3417354..3417959	3417591	606	5.79E-04	3418038..3418550	3418363	513	7.32E-03
c5164	0	Intergenic	4931257..4931885	4931651	629	6.81E-04	4931424..4931959	4931613	536	3.00E-02
ygiK	0	Intragenic	3588884..3589367	3589118	484	1.00E-03	3588918..3589335	3589147	418	5.00E-02
yfaL	0	Intergenic	2630987..2631481	2631245	495	1.24E-03	2631013..2631424	2631202	412	2.31E-04
c2406	-190	Intergenic	2209958..2210494	2210195	537	1.32E-03	2210043..2210457	2210269	415	8.27E-03

ygiL	0	Intergenic	3624039..3624631	3624395	593	1.35E-03	3624167..3624606	3624418	440	9.37E-04
fimE	0	Intergenic	5136286..5136775	5136521	490	1.52E-03	5136363..5136759	5136552	397	4.00E-02
envR	0	Intergenic	3845140..3845648	3845414	509	2.14E-03	3845218..3845607	3845419	390	2.39E-03
c4896	0	Intragenic	4655944..4656436	4656181	493	2.15E-03	4656051..4656439	4656251	389	7.32E-03
c2348	0	Intergenic	2160258..2160757	2160493	500	2.21E-03	2160279..2160658	2160468	380	3.06E-04
yadN	0	Intergenic	167376..167865	167613	490	2.31E-03	167517..167920	167732	404	2.14E-04
yjdA	0	Intergenic	4885292..4885787	4885551	496	2.57E-03	4885358..4885754	4885547	397	4.89E-03
fucP	0	Intragenic	3208553..3209172	3208788	620	2.65E-03	3208638..3209130	3208827	493	4.00E-02
nanC	0	Intergenic	5134074..5134599	5134309	526	3.38E-03	5134134..5134528	5134341	395	2.00E-02
ydbA_1	0	Intergenic	1663930..1664429	1664195	500	3.91E-03	1663913..1664348	1664160	436	7.37E-03
ompC	0	Intergenic	1421035..1421524	1421271	490	4.73E-03	1421060..1421453	1421248	394	6.79E-03
bglG	0	Intergenic	4407383..4408063	4407827	681	4.81E-03	4407738..4408118	4407932	381	3.00E-02
c5298	0	Intergenic	5038053..5038523	5038288	471	4.84E-03	5038057..5038439	5038246	383	3.87E-03
yicO	0	Intragenic	4348267..4348913	4348677	647	5.01E-03	4348556..4348934	4348745	379	3.00E-02
frc	0	Intergenic	2775657..2776157	2775924	501	5.94E-03	2775616..2776057	2775805	442	6.82E-03
c1618	0	Intragenic	1460905..1461431	1461195	527	6.11E-03	1460930..1461353	1461119	424	3.00E-02
c0407	0	Intragenic	388028..388586	388262	559	6.37E-03	387986..388442	388254	457	5.00E-02
c4894	0	Intergenic	4652769..4653327	4653004	559	6.40E-03	4652686..4653084	4652897	399	2.00E-02
sat	0	Intergenic	3460036..3460533	3460297	498	6.68E-03	3460091..3460470	3460282	380	1.00E-02
pitB	0	Intergenic	3558084..3558565	3558331	482	6.90E-03	3558160..3558572	3558384	413	1.00E-02
tnaA	0	Intergenic	4390079..4390574	4390316	496	1.00E-02	4390047..4390517	4390329	471	3.00E-02
tdcB	0	Intergenic	3705304..3705792	3705541	489	1.00E-02	3705354..3705737	3705549	384	3.00E-02
ypdI	0	Intergenic	2776873..2777347	2777110	475	1.00E-02	2776854..2777242	2777054	389	2.00E-02
c5088	0	Intragenic	4863274..4863764	4863528	491	1.00E-02	4863446..4863833	4863635	388	2.00E-02
c4492	0	Intergenic	4275822..4276303	4276067	482	1.00E-02	4275872..4276340	4276061	469	2.10E-03
c2815	0	Intergenic	2670640..2671153	2670875	514	1.00E-02	2670713..2671119	2670902	407	5.76E-03
iraD	18	Intergenic	5149972..5150464	5150207	493	1.00E-02	5149969..5150413	5150225	445	1.00E-02

c3712	0	Intergenic	3544866..3545334	3545101	469	1.00E-02	3544928..3545394	3545206	467	2.00E-02
yegR	0	Intergenic	2459372..2459884	2459607	513	1.00E-02	2459456..2459853	2459643	398	2.00E-02
c4020	0	Intergenic	3836796..3837307	3837031	512	2.00E-02	3836896..3837303	3837085	408	1.00E-02
yebN	-133	Intergenic	2056751..2057226	2056988	476	2.00E-02	2056859..2057294	2057048	436	2.00E-02
yliE	483	Intragenic	894352..894882	894648	531	2.00E-02	894394..894792	894604	399	3.00E-02
c3031	0	Intergenic	2903888..2904394	2904122	507	2.00E-02	2903942..2904350	2904162	409	1.62E-03
yjcS	0	Intergenic	4865626..4866130	4865863	505	2.00E-02	4865608..4866002	4865815	395	2.00E-02
dsdC	0	Intergenic	2759612..2760084	2759848	473	3.00E-02	2759962..2760486	2760151	525	9.03E-04
stpA	0	Intergenic	3076982..3077460	3077219	479	3.00E-02	3077052..3077532	3077241	481	4.00E-02
yeeN	0	Intergenic	2256940..2257468	2257177	529	4.00E-02	2256986..2257371	2257175	386	2.00E-02
yheE	0	Intragenic	3886779..3887278	3887016	500	5.00E-02	3886745..3887144	3886956	400	4.00E-02
fdeC	0	Intergenic	395223..395931	395460	709	5.00E-02	395288..395692	395477	405	2.00E-02

Table S-2 CFT073 binding sites (+ D-ser)

			Replicate 1				Replicate 2			
Gene	**Dist. Gene**	**Context**	**Co-ordinates**	**Centre**	**Length**	***p*-value**	**Co-ordinates**	**Centre**	**Length**	***p*-value**
rfaJ	0	Intragenic	4234886..4235313	4235116	428	1.83E-15	4234953..4235367	4235155	415	6.13E-19
c4214	-88	Intergenic	3999713..4000110	3999911	398	1.31E-12	3999692..4000097	3999896	406	5.78E-10
waaL	511	Intragenic	4230330..4230726	4230528	397	2.78E-11	4230325..4230760	4230526	436	3.29E-16
waaV	0	Intergenic	4232110..4232540	4232343	431	6.50E-11	4232127..4232555	4232357	429	1.91E-14
yhhZ	0	Intergenic	4021047..4021474	4021245	428	3.42E-10	4020894..4021491	4021096	598	1.22E-10
ygiL	0	Intergenic	3624118..3624613	3624416	496	5.70E-10	3624129..3624532	3624332	404	6.51E-08
yadN	-44	Intergenic	167507..167907	167705	401	7.12E-10	167478..167898	167697	421	2.25E-10

c0080	0	Intragenic	74758..75189	74955	432	2.39E-09	74949..75356	75151	408	1.88E-10
c3694	0	Intragenic	3525853..3526314	3526117	462	2.54E-09	3525897..3526325	3526099	429	3.43E-10
c4739	0	Intergenic	4507176..4507645	4507448	470	3.87E-09	4507165..4507576	4507375	412	4.87E-12
c2897	-52	Intergenic	2757411..2757939	2757609	529	2.96E-08	2757468..2757891	2757692	424	8.94E-11
c4498	847	Intergenic	4282776..4283226	4282974	451	4.43E-08	4282810..4283273	4283012	464	4.17E-07
c5435	0	Intergenic	5181703..5182135	5181901	433	4.45E-08	5181701..5182152	5181951	452	3.37E-07
c3692	776	Intergenic	3519957..3520396	3520155	440	5.56E-08	3519956..3520336	3520158	381	1.83E-06
c4896	0	Intergenic	4656052..4656453	4656256	402	6.68E-08	4656015..4656455	4656254	441	6.45E-08
yfcV	-33	Intergenic	2737755..2738225	2738028	471	7.76E-08	2737762..2738198	2737998	437	4.27E-06
c3686	-144	intergenic	3511546..3511947	3511750	402	1.29E-07	3511502..3511925	3511724	424	1.00E-09
c4492	-71	Intergenic	4275814..4276225	4276012	412	2.28E-07	4275811..4276260	4276059	450	3.24E-11
asst	0	Intergenic	3618401..3618899	3618702	499	2.80E-07	3618370..3618840	3618639	471	2.75E-08
leuO	0	Intergenic	87703..88224	87901	522	2.82E-07	87736..88181	87938	446	2.37E-06
cmtB	0	Intergenic	3373110..3373525	3373328	416	3.62E-07	3373088..3373543	3373342	456	2.96E-07
c0322	0	Intergenic	292319..292758	292517	440	3.99E-07	292337..292791	292539	455	1.31E-07
ycdT	0	Intergenic	1125690..1126096	1125900	407	4.74E-07	1125693..1126125	1125895	433	5.07E-06
c4423	0	Intergenic	4204935..4205359	4205162	425	7.60E-07	4204879..4205363	4205162	485	8.04E-07
rfaY	0	Intragenic	4233877..4234333	4234073	457	7.70E-07	4233832..4234313	4234034	482	5.15E-13
papA	0	Intergenic	3438020..3438517	3438320	498	1.13E-06	3438032..3438492	3438291	461	3.14E-08
yfaL	0	Intergenic	2631023..2631424	2631221	402	1.17E-06	2630979..2631459	2631181	481	4.47E-07
c0327	-22	Intergenic	299139..299565	299337	427	1.32E-06	299083..299498	299297	416	1.68E-07
caiT	0	Intergenic	44502..45047	44700	546	1.58E-06	44497..45045	44699	549	2.67E-07
yjjQ	-27	Intergenic	5191990..5192438	5192188	449	1.85E-06	5191968..5192375	5192174	408	7.68E-07
fimE	-15	Intergenic	5136346..5136753	5136544	408	2.22E-06	5136282..5136722	5136521	441	5.52E-10
pitB	0	Intergenic	3558153..3558579	3558351	427	2.31E-06	3558159..3558582	3558381	424	8.80E-07
c0077	0	Intergenic	73941..74415	74218	475	2.50E-06	73872..74426	74225	555	6.85E-08
c5298	0	Intergenic	5038049..5038452	5038255	404	2.54E-06	5038053..5038523	5038255	471	1.82E-11

c5209	-157	Intragenic	4964202..4964678	4964481	477	2.63E-06	4963981..4964586	4964183	606	1.88E-08
yhiM	0	Intergenic	4068821..4069290	4069019	470	2.83E-06	4068879..4069312	4069111	434	1.03E-10
pagP	-67	Intergenic	696667..697080	696883	414	3.48E-06	696605..697166	696807	562	4.66E-08
hlyC	-486	Intergenic	3417322..3417723	3417520	402	3.64E-06	3418168..3418696	3418370	529	8.40E-06
envR	0	Intergenic	3845196..3845596	3845394	401	3.66E-06	3845204..3845621	3845420	418	1.47E-09
c0325	822	Intragenic	297396..297797	297600	402	3.81E-06	297399..297913	297600	515	5.78E-08
c3031	0	Intergenic	2903918..2904358	2904162	441	3.95E-06	2903936..2904363	2904162	428	3.09E-08
c4303	0	Intergenic	4082825..4083321	4083023	497	4.56E-06	4082875..4083342	4083077	468	1.20E-08
c5426	182	Intragenic	5171570..5171978	5171768	409	4.74E-06	5171408..5171969	5171769	562	1.20E-06
garP	0	Intergenic	3712121..3712663	3712319	543	5.12E-06	3712191..3712685	3712391	495	3.29E-06
cydA	-944	Intergenic	792774..793183	792986	410	5.20E-06	792775..793255	792977	481	2.57E-04
yjdA	0	Intergenic	4885306..4885790	4885504	485	6.20E-06	4885309..4885732	4885511	424	1.15E-08
tauA	-85	Intergenic	460039..460438	460241	400	6.69E-06	459988..460394	460190	407	8.09E-09
tnaA	0	Intragenic	4390031..4390481	4390229	451	7.73E-06	4390075..4390547	4390346	473	3.51E-08
c3766	0	Intragenic	3600368..3600870	3600566	503	9.08E-06	3600466..3600989	3600668	524	6.70E-07
c4994	0	Intergenic	4771197..4771598	4771395	402	1.07E-05	4771190..4771600	4771399	411	4.69E-07
c0430	-178	Intergenic	415201..415788	415591	588	1.20E-05	415235..415742	415541	508	1.66E-06
c3302	0	Intragenic	3148555..3148969	3148753	415	1.33E-05	3148531..3148975	3148733	445	6.04E-07
arcA	-39	Intergenic	5094995..5095446	5095193	452	1.48E-05	5094903..5095360	5095159	458	2.63E-07
frc	0	Intergenic	2775625..2776068	2775823	444	1.64E-05	2775657..2776110	2775909	454	1.18E-06
yiiE	0	Intergenic	4603113..4603511	4603311	399	1.86E-05	4603087..4603490	4603289	404	4.56E-06
yghT	0	Intergenic	3555739..3556173	3555977	435	2.15E-05	3555714..3556137	3555936	424	6.30E-06
wzy	-129	Intragenic	2399205..2399624	2399403	420	2.18E-05	2399200..2399653	2399402	454	6.20E-06
dpiB	0	Intergenic	692480..692916	692719	437	2.55E-05	692470..693001	692671	532	1.15E-05
ompC	-1	Intergenic	1421091..1421510	1421313	420	2.60E-05	1421062..1421481	1421280	420	2.08E-07
tssB	-39	Intergenic	3222559..3222962	3222757	404	2.74E-05	3222592..3223034	3222794	443	7.29E-06
c4920	0	Intergenic	4685058..4685508	4685256	451	3.40E-05	4685058..4685482	4685260	425	1.31E-04

yiaY	0	Intergenic	4193450..4193849	4193652	400	3.46E-05	4193398..4193822	4193621	425	4.56E-09
fdeC	0	Intergenic	395279..395770	395477	492	3.53E-05	395267..395775	395467	509	3.51E-04
gadA	0	Intergenic	4106198..4106634	4106394	437	3.63E-05	4106162..4106645	4106364	484	1.94E-06
c3712	0	Intergenic	3544937..3545337	3545140	401	3.93E-05	3544890..3545332	3545131	443	5.09E-07
tdcB	0	Intergenic	3705272..3705738	3705541	467	5.65E-05	3705256..3705718	3705524	463	7.11E-06
c0346	0	Intragenic	322315..322739	322542	425	5.78E-05	322298..322742	322542	445	3.12E-04
sat	0	Intergenic	3460004..3460473	3460276	470	5.88E-05	3460045..3460516	3460315	472	1.39E-07
c1271	140	Intragenic	1222173..1222602	1222405	430	6.27E-05	1221994..1222596	1222396	603	7.71E-06
hemB	-16	Intergenic	465002..465419	465200	418	6.50E-05	465037..465525	465239	489	1.51E-06
c1208	-75	Intragenic	1164820..1165259	1165062	440	7.39E-05	1164817..1165259	1165019	443	2.46E-05
ypdI	0	Intergenic	2776802..2777225	2777028	424	9.36E-05	2776782..2777206	2777006	425	1.10E-04
c4020	0	Intergenic	3836892..3837293	3837096	402	9.76E-05	3836949..3837363	3837150	415	3.63E-05
wzx	0	Intragenic	2399822..2400397	2400200	576	1.14E-04	2399804..2400354	2400153	551	1.73E-05
nanC	0	Intergenic	5134093..5134568	5134291	476	1.26E-04	5134148..5134623	5134350	476	1.18E-07
bglG	0	Intergenic	4407620..4408101	4407904	482	1.33E-04	4407663..4408091	4407890	429	3.94E-05
livK	0	Intergenic	4036062..4036463	4036266	402	1.35E-04	4036072..4036507	4036306	436	4.65E-07
hcp	0	Intragenic	4521230..4521804	4521608	575	1.37E-04	4521256..4521815	4521614	560	1.43E-07
c5088	0	Intergenic	4863464..4863897	4863700	434	1.37E-04	4863425..4863862	4863661	438	6.86E-06
yjeJ	0	Intergenic	4982924..4983378	4983182	455	1.48E-04	4982892..4983346	4983147	455	2.94E-06
c2383	0	Intergenic	2185351..2185883	2185686	533	1.59E-04	2185507..2185921	2185709	415	6.37E-06
agaB	0	Intergenic	3723469..3723879	3723667	411	2.21E-04	3723574..3724084	3723776	511	4.65E-05
c4488	0	Intergenic	4269841..4270240	4270039	400	2.30E-04	4269762..4270366	4270165	605	6.08E-05
yicO	0	Intergenic	4348466..4348882	4348685	417	2.33E-04	4348477..4348891	4348690	415	4.06E-07
yebN	-137	Intergenic	2056792..2057222	2057025	431	2.47E-04	2056808..2057244	2057043	437	1.04E-05
c2348	0	Intergenic	2160276..2160721	2160474	446	2.94E-04	2160271..2160721	2160473	451	4.35E-06
adiY	0	Intergenic	4896255..4896822	4896625	568	3.06E-04	4896371..4896845	4896644	475	5.31E-05
yheE	0	Intergenic	3886738..3887177	3886936	440	3.21E-04	3886740..3887182	3886942	443	1.31E-03

c2815	0	Intergenic	2670706..2671101	2670903	396	3.36E-04	2670699..2671213	2670901	515	3.62E-04
fimA	-43	Intergenic	5137381..5137802	5137578	422	3.47E-04	5137377..5137833	5137632	457	1.34E-06
aslA	0	Intergenic	4486185..4486610	4486413	426	3.53E-04	4486170..4486587	4486372	418	4.94E-05
fucP	0	Intergenic	3208564..3209130	3208762	567	4.02E-04	3208613..3209029	3208828	417	1.26E-05
yliE	525	Intragenic	894436..894847	894650	412	4.11E-04	894428..894834	894630	407	6.69E-05
yjcS	0	Intergenic	4865663..4866083	4865861	421	4.43E-04	4865638..4866047	4865840	410	1.04E-05
yegR	0	Intergenic	2459395..2459841	2459644	447	4.46E-04	2459427..2459835	2459629	409	1.61E-06
c4983	0	Intragenic	4762924..4763345	4763122	422	4.56E-04	4762918..4763329	4763129	412	5.62E-04
c5356	0	Intergenic	5098404..5098854	5098657	451	5.61E-04	5098440..5098906	5098642	467	2.92E-05
c1166	0	Intragenic	1130647..1131048	1130845	402	5.75E-04	1130507..1131038	1130837	532	9.96E-04
chiA	0	Intergenic	3901180..3901576	3901379	397	5.93E-04	3901085..3901616	3901287	532	1.38E-03
c0407	-249	Intergenic	388135..388577	388333	443	6.27E-04	388091..388542	388293	452	4.87E-06
yhbX	0	Intergenic	3755096..3755491	3755294	396	6.38E-04	3755062..3755519	3755318	458	9.79E-05
c1892	0	Intragenic	1726230..1726742	1726428	513	7.00E-04	1726250..1726720	1726451	471	4.17E-04
c1618	0	Intragenic	1460993..1461393	1461197	401	7.14E-04	1461015..1461465	1461217	451	1.50E-04
yebB	0	Intragenic	2097660..2098115	2097858	456	7.40E-04	2097660..2098263	2097861	604	3.19E-03
ydbA_1	0	Intergenic	1663883..1664336	1664139	454	7.72E-04	1663929..1664335	1664134	407	1.82E-03
ykgI	0	Intragenic	402822..403248	403051	427	8.51E-04	402817..403255	403054	439	1.23E-04
yhfL	0	Intergenic	3930528..3930927	3930725	400	1.14E-03	3930598..3931017	3930816	420	1.83E-05
dsdC	0	Intergenic	2759996..2760586	2760194	591	1.15E-03	2759949..2760549	2760151	601	6.20E-07
yeeN	0	Intergenic	2257022..2257582	2257220	561	1.40E-03	2257054..2257516	2257256	463	2.03E-05
c4986	0	Intergenic	4764725..4765138	4764941	414	1.49E-03	4764689..4765105	4764889	417	3.10E-04
c4502	0	Intergenic	4289649..4290046	4289849	398	1.55E-03	4289593..4290035	4289834	443	9.44E-04
stpA	0	Intergenic	3077102..3077562	3077300	461	1.65E-03	3077092..3077555	3077294	464	9.26E-06
c2406	-296	Intergenic	2210064..2210461	2210262	398	1.78E-03	2210618..2211182	2210981	565	8.23E-06
fdrA	0	Intergenic	615034..615477	615232	444	2.19E-03	615068..615503	615302	436	1.20E-04
cyoA	-126	Intergenic	525665..526079	525882	415	2.22E-03	525588..525993	525790	406	9.98E-04

c4015	0	Intragenic	3833074..3833654	3833457	581	2.41E-03	3833248..3833688	3833489	441	1.52E-04
rpiB	0	Intergenic	4872115..4872551	4872313	437	2.55E-03	4872035..4872514	4872317	480	1.89E-03
iraP	0	Intergenic	475076..475575	475274	500	3.11E-03	475090..475506	475305	417	2.72E-06
iraD	-36	Intergenic	5149995..5150446	5150192	452	3.14E-03	5149971..5150394	5150193	424	3.70E-04
ompN	0	Intergenic	1654547..1654942	1654745	396	3.36E-03	1654537..1655003	1654802	467	4.80E-03
c4539	0	Intergenic	4318776..4319205	4319008	430	4.57E-03	4318823..4319242	4319024	420	1.00E-02
yffB	0	Intergenic	2861671..2862106	2861909	436	5.69E-03	2861706..2862127	2861926	422	1.16E-03
yciF	0	Intergenic	1559655..1560054	1559853	400	5.84E-03	1559522..1559938	1559737	417	1.19E-03
araE	0	Intergenic	3290174..3290572	3290372	399	6.63E-03	3290122..3290525	3290324	404	6.48E-03
ais	0	Intergenic	2651820..2652398	2652201	579	7.24E-03	2651952..2652392	2652191	441	2.71E-04
c0363	0	Intergenic	340655..341100	340903	446	8.01E-03	340627..341037	340829	411	2.39E-03
bdm	0	Intergenic	1749132..1749601	1749404	470	8.76E-03	1749199..1749617	1749416	419	1.98E-03
c0410	0	Intergenic	390881..391284	391079	404	2.00E-02	390857..391392	391048	536	1.00E-02
clbR	-158	Intergenic	2313770..2314181	2313984	412	3.00E-02	2314636..2315118	2314831	483	4.01E-03

Table S-3 CE10 DsdC1 binding data (no D-ser).

			Replicate 1				Replicate 2			
Gene	**Dist. Gene**	**Context**	**Co-ordinates**	**Centre**	**Length**	***p*-value**	**Co-ordinates**	**Centre**	**Length**	***p*-value**
CE10_0023	-701	Intragenic	27441..27818	27655	378	0.00+00E	27570..27709	27655	140	6.54E-45
CE10_0067	0	Intragenic	79557..79975	79725	419	0.00+00E	79539..79698	79594	160	5.26E-17
yadN	0	Intergenic	166926..167348	167185	423	0.00+00E	166217..166355	166272	139	6.19E-12
tauA	-100	Intergenic	365148..365547	365378	400	0.00+00E	365288..365432	365378	145	3.50E-23
CE10_0889	0	Intergenic	940940..941277	941109	338	0.00+00E	941073..941226	941173	154	6.93E-17

ysdS	0	Intergenic	2249533..2249989	2249820	457	0.00+00E	2249802..2249931	2249877	130	1.99E-18
wbbC	0	Intragenic	2367786..2368225	2368057	440	0.00+00E	2367937..2368062	2368008	126	7.20E-71
wzy	-105	Intergenic	2368985..2369359	2369154	375	0.00+00E	2369039..2369172	2369118	134	8.65E-37
CE10_2740	4	Intergenic	2788664..2789046	2788830	383	0.00+00E	2788877..2788999	2788949	123	1.66E-31
dsdC1	0	Intergenic	2791235..2791582	2791414	348	0.00+00E	2791368..2791477	2791423	110	9.87E-156
CE10_3076	0	Intergenic	3138070..3138508	3138240	439	0.00+00E	3138181..3138283	3138234	103	9.80E-26
yqeI	-59	Intergenic	3340403..3340735	3340557	333	0.00+00E	3340524..3340674	3340624	151	4.97E-33
ygeG	-62	Intergenic	3343448..3343743	3343556	296	0.00+00E	3343481..3343606	3343533	126	3.75E-43
prgI	28	Intragenic	3350452..3350809	3350656	358	0.00+00E	3350640..3350776	3350722	137	1.71E-28
eprH	0	Intergenic	3351564..3351974	3351806	411	0.00+00E	3351640..3351757	3351705	118	4.76E-34
epaS	0	Intragenic	3354052..3354436	3354278	385	0.00+00E	3354192..3354316	3354262	125	2.86E-46
epaO	0	Intergenic	3356511..3356931	3356764	421	0.00+00E	3356785..3356857	3356814	73	1.17E-20
neuE	-169	Intragenic	3544618..3544986	3544780	369	0.00+00E	3544725..3544861	3544780	137	7.61E-24
neuA	-364	Intragenic	3547233..3547655	3547381	423	0.00+00E	3546798..3546956	3546853	159	4.32E-14
neuB	-278	Intergenic	3548187..3548660	3548503	474	0.00+00E	3548362..3548475	3548417	114	9.15E-29
slp	-296	Intragenic	4101919..4102392	4102088	474	0.00+00E	4102402..4102530	4102477	129	2.34E-29
rfaL	0	Intragenic	4258249..4258601	4258438	353	0.00+00E	4258330..4258481	4258436	152	3.80E-26
waaV	0	Intergenic	4260038..4260392	4260235	355	0.00+00E	4260181..4260334	4260235	154	1.76E-29
waaT	0	Intergenic	4262851..4263201	4263019	351	0.00+00E	4262962..4263125	4263017	164	7.34E-31
dsdC2	0	Intergenic	4332276..4332616	4332446	341	0.00+00E	4332401..4332520	4332456	120	7.67E-151
sipD	0	Intergenic	4365627..4366074	4365905	448	0.00+00E	4365896..4366048	4365996	153	2.65E-18
CE10_4297	0	Intergenic	4367298..4367739	4367571	442	0.00+00E	4367521..4367649	4367595	129	7.93E-24
espY4	0	Intragenic	4441550..4442034	4441719	485	0.00+00E	4441179..4441292	4441238	114	2.01E-19
CE10_4362	118	Intragenic	4442253..4442560	4442395	308	0.00+00E	4442079..4442190	4442136	112	1.52E-29
CE10_4466	-357	Intragenic	4553284..4553676	4553453	393	0.00+00E	4552933..4553055	4553001	123	1.16E-09
espX4	0	Intragenic	4852614..4853038	4852784	425	0.00+00E	4852729..4852847	4852784	119	1.33E-26

ubiC	-671	Intragenic	4853255..4853725	4853565	471	0.00+00E	4853445..4853536	4853500	92	3.11E-32
yjbM	0	Intragenic	4862216..4862614	4862386	399	0.00+00E	4863124..4863266	4863178	143	2.15E-15
CE10_4768	163	Intragenic	4870355..4870783	4870524	429	0.00+00E	4870388..4870551	4870443	164	3.25E-26
CE10_4943	0	Intergenic	5047562..5047946	5047777	385	0.00+00E	5047680..5047795	5047735	116	5.73E-34
fimE	-48	Intergenic	5174559..5174954	5174728	396	0.00+00E	5174609..5174720	5174663	112	1.45E-08
yjiC	0	Intergenic	5189283..5189779	5189610	497	0.00+00E	5189546..5189702	5189653	157	3.65E-27
yhhZ	0	Intragenic	4029255..4029623	4029454	369	6.40E-322	4029061..4029192	4029138	132	2.26E-29
yfaL	0	Intergenic	2651239..2651608	2651399	370	8.47E-318	2650745..2650878	2650824	134	6.18E-13
eivC	-375	Intragenic	3360277..3360778	3360445	502	2.69E-314	3360412..3360544	3360493	133	8.52E-23
hutU	0	Intergenic	769533..769906	769737	374	2.62E-312	769648..769781	769727	134	2.00E-14
dusA	-266	Intragenic	4862853..4863311	4863013	459	1.23E-307	4862252..4862408	4862306	157	3.57E-33
sipB	0	Intergenic	4362509..4363016	4362847	508	1.02E-306	4362820..4362943	4362875	124	1.88E-25
CE10_4870	0	Intergenic	4974868..4975249	4975038	382	7.45E-303	4863124..4863266	4863178	143	2.15E-15
rfaY	0	Intragenic	4261765..4262267	4261929	503	4.35E-293	4974995..4975105	4975051	111	2.72E-38
fucA	-12	Intergenic	3285996..3286441	3286164	446	3.97E-291	3286095..3286228	3286149	134	1.41E-21
leuO	-90	Intergenic	92721..93073	92888	353	5.01E-291	92834..92931	92887	98	2.96E-16
yghJ	0	Intergenic	3567445..3567945	3567615	501	3.00E-283	3567538..3567665	3567579	128	7.89E-16
cmtB	0	Intragenic	3449900..3450402	3450233	503	1.08E-270	3450052..3450182	3450106	131	9.36E-16
ydeK	-49	Intergenic	1743092..1743461	1743292	370	9.24E-265	1743217..1743352	1743298	136	1.71E-19
CE10_2939	0	Intergenic	2996411..2996816	2996581	406	2.17E-264	2996566..2996705	2996651	140	1.34E-18
envR	0	Intergenic	3874440..3874808	3874610	369	6.73E-263	3874519..3874636	3874566	118	5.87E-21
caiT	-60	Intergenic	50204..50678	50509	475	1.12E-260	50442..50571	50517	130	1.62E-14
CE10_4228	-73	Intergenic	4305007..4305391	4305176	385	7.04E-259	4305108..4305261	4305163	154	7.41E-09
sfmA	0	Intergenic	561180..561551	561349	372	9.25E-253	561286..561403	561349	118	4.50E-31
yliE	-718	Intragenic	908923..909268	909093	346	2.83E-243	909011..909154	909066	144	3.70E-16
CE10_3627	0	Intergenic	3707366..3707869	3707701	504	2.06E-242	3707646..3707779	3707736	134	1.38E-13

yhfL	0	Intergenic	3944853..3945223	3945023	371	2.62E-240	3944945..3945046	3945001	102	1.47E-17
ycbQ	0	Intergenic	1039138..1039516	1039308	379	2.65E-235	1039278..1039383	1039330	106	2.83E-13
kpsM	-233	Intragenic	3550328..3550769	3550600	442	4.32E-233	3550189..3550321	3550267	133	7.16E-13
yjbE	0	Intergenic	4839210..4839675	4839506	466	1.67E-230	4839462..4839591	4839493	130	1.79E-15
CE10_1660	299	Intragenic	1690807..1691181	1691023	375	1.10E-227	1690990..1691128	1691045	139	1.47E-09
CE10_1948	-40	Intergenic	1970983..1971368	1971152	386	3.21E-212	1971119..1971281	1971227	163	1.67E-07
yeeN	0	Intergenic	2298004..2298367	2298203	364	6.41E-204	2298099..2298252	2298203	154	1.92E-10
CE10_4567	-60	Intergenic	4657507..4657867	4657703	361	1.60E-203	4657648..4657784	4657701	137	4.46E-18
CE10_4598	0	Intergenic	4685814..4686203	4686034	390	4.41E-200	4685991..4686151	4686046	161	7.97E-21
setC	-19	Intergenic	4358033..4358399	4358240	367	9.56E-199	4358168..4358290	4358216	123	6.06E-14
CE10_0066	0	Intergenic	78738..79150	78986	413	4.80E-195	78834..78948	78896	115	7.89E-17
yiiG	0	Intergenic	4651369..4651762	4651539	394	7.58E-194	4651497..4651615	4651561	119	7.52E-27
csiD	-303	Intragenic	3140284..3140638	3140454	355	3.30E-184	3140404..3140518	3140458	115	6.38E-21
stpA	0	Intergenic	3150387..3150730	3150557	344	5.20E-184	3150497..3150637	3150551	141	1.18E-12
CE10_1661	0	Intergenic	1692605..1692994	1692825	390	1.37E-181	1692767..1692883	1692821	117	7.50E-15
yiaW	0	Intergenic	4217589..4217945	4217759	357	2.10E-166	4217675..4217805	4217751	131	4.27E-12
ecpD	0	Intergenic	166102..166449	166272	348	1.22E-165	166217..166355	166272	139	6.19E-12
fimB	0	Intergenic	5173511..5173987	5173681	477	1.33E-164	5173596..5173706	5173649	111	3.52E-05
speF	-187	Intergenic	730553..731056	730887	504	1.88E-164	730743..730811	730784	69	1.12E-11
CE10_5175	999	Intragenic	5280007..5280442	5280177	436	3.62E-163	5280928..5281088	5281034	161	1.08E-26
fucP	0	Intragenic	3286583..3286944	3286776	362	2.65E-162	3286722..3286845	3286777	124	2.52E-23
yiaY	0	Intergenic	4222357..4222771	4222602	415	9.83E-157	4222623..4222735	4222681	113	5.49E-13
fdrA	0	Intergenic	549560..550008	549839	449	7.72E-156	549686..549819	549741	134	8.66E-16
yqiK	0	Intergenic	3647776..3648120	3647951	345	3.20E-155	3647855..3647992	3647938	138	1.81E-10
ais	0	Intergenic	2677305..2677651	2677475	347	1.35E-151	2677414..2677528	2677464	115	3.48E-16
yfbL	0	Intergenic	2697398..2697781	2697568	384	5.01E-151	2697489..2697584	2697547	96	1.13E-11

ypdI	0	Intergenic	2808139..2808484	2808306	346	2.05E-147	2808225..2808359	2808279	135	7.36E-09
nanC	0	Intergenic	5172391..5172767	5172598	377	2.72E-144	5172451..5172570	5172517	120	4.46E-12
rpmE2	-210	Intergenic	281778..282115	281951	338	6.69E-143	281850..282013	281905	164	1.19E-08
yeiT	0	Intergenic	2539340..2539736	2539510	397	4.77E-137	2539468..2539591	2539520	124	5.37E-13
yobF	-202	Intergenic	2126724..2127140	2126973	417	1.52E-135	2126925..2127080	2127028	156	2.51E-12
yahL	0	Intergenic	326755..327210	326924	456	2.96E-127	326904..327028	326959	125	4.71E-11
CE10_0335	-24	Intergenic	370032..370485	370193	454	3.63E-121	370106..370265	370211	160	2.08E-09
CE10_3982	0	Intragenic	4045074..4045400	4045219	327	1.26E-120	4045141..4045246	4045203	106	1.91E-11
agaB	0	Intergenic	3746884..3747262	3747100	379	9.08E-109	3747000..3747149	3747095	150	3.80E-16
neuO	59	Intragenic	839634..840140	839803	507	1.98E-106	839932..840043	839987	112	4.87E-07
ycbR	0	Intergenic	1039827..1040283	1039995	457	1.61E-103	1039972..1040091	1040027	120	3.18E-18
yjeJ	0	Intergenic	4993586..4993924	4993757	339	1.01E-101	4993712..4993820	4993766	109	6.35E-11

Table S-4 CE10 DsdC1 binding sites (+ D-ser).

			Replicate 1				Replicate 2			
Gene	**Dist. Gene**	**Context**	**Co-ordinates**	**Centre**	**Length**	***p*-value**	**Co-ordinates**	**Centre**	**Length**	***p*-value**
CE10_0023	-671	Intragenic	27499..27746	27649	248	0.00E+00	27439..27831	27664	393	7.89E-127
tauA	-160	Intergenic	365260..365487	365354	228	0.00E+00	365165..365532	365363	368	2.09E-98
wbbC	0	Intragenic	2367910..2368130	2368008	221	0.00E+00	2367780..2368239	2368070	460	3.24E-141
CE10_2740	-203	Intergenic	2788863..2789052	2788948	190	0.00E+00	2789978..2790338	2790146	361	3.88E-100
dsdC1	-5	Intergenic	2791325..2791520	2791423	196	0.00E+00	2791243..2791582	2791413	340	3.73E-196
CE10_2939	0	Intergenic	2996531..2996749	2996652	219	0.00E+00	2996439..2996793	2996624	355	2.76E-77
ygeG	0	Intergenic	3343554..3343783	3343652	230	0.00E+00	3342998..3343347	3343186	350	5.89E-100
CE10_3282	0	Intergenic	3347661..3347860	3347759	200	0.00E+00	3347572..3348074	3347740	503	1.92E-126

eprH	-55	Intergenic	3351683..3351896	3351799	214	0.00E+00	3351566..3351932	3351763	367	2.50E-88
neuB	-378	Intergenic	3548287..3548559	3548384	273	0.00E+00	3548221..3548619	3548480	399	1.97E-111
yhhZ	0	Intergenic	4029031..4029235	4029138	205	0.00E+00	4029334..4029732	4029497	399	2.42E-77
yiaW	0	Intergenic	4217648..4217872	4217775	225	0.00E+00	4217597..4217952	4217784	356	4.65E-42
CE10_4160	0	Intragenic	4233642..4233911	4233814	270	0.00E+00	4233461..4233943	4233774	483	3.25E-97
rfaL	615	Intragenic	4258299..4258534	4258437	236	0.00E+00	4258248..4258605	4258436	358	2.53E-115
waaV	0	Intergenic	4260105..4260377	4260280	273	0.00E+00	4260039..4260378	4260209	340	6.41E-100
waaT	0	Intergenic	4262918..4263149	4263016	232	0.00E+00	4262799..4263173	4263006	375	1.34E-121
CE10_4229	0	Intergenic	4308317..4308519	4308422	203	0.00E+00	4308239..4308607	4308409	369	1.28E-72
dsdC2	0	Intergenic	4332349..4332573	4332447	225	0.00E+00	4332272..4332621	4332452	350	5.07E-201
CE10_4270	0	Intragenic	4347338..4347592	4347497	255	0.00E+00	4347255..4347740	4347425	486	7.41E-121
sipD	0	Intergenic	4365860..4366057	4365960	198	0.00E+00	4365603..4366105	4365937	503	4.38E-86
CE10_4297	0	Intergenic	4367470..4367709	4367612	240	0.00E+00	4368027..4368395	4368231	369	5.29E-74
yjbE	-61	Intergenic	4839351..4839589	4839492	239	0.00E+00	4839261..4839654	4839430	394	9.42E-73
espX4	0	Intragenic	4852657..4852854	4852754	198	0.00E+00	4852605..4853086	4852775	482	8.16E-99
ubiC	0	Intragenic	4853401..4853639	4853499	239	0.00E+00	4853330..4853734	4853565	405	6.39E-119
yjbM	0	Intragenic	4862291..4862525	4862389	235	0.00E+00	4862208..4862551	4862375	344	8.68E-96
CE10_4768	317	Intragenic	4870344..4870630	4870442	287	0.00E+00	4870317..4870811	4870487	495	2.86E-98
CE10_4870	0	Intergenic	4974954..4975169	4975072	216	0.00E+00	4974872..4975235	4975066	364	2.33E-75
CE10_4943	0	Intergenic	5047640..5047864	5047738	225	0.00E+00	5047570..5047909	5047740	340	2.50E-93
yjiC	-179	Intergenic	5189472..5189709	5189613	238	0.00E+00	5189325..5189740	5189571	416	1.33E-108
epaS	265	Intragenic	3354076..3354339	3354242	264	3.10E-322	3354022..3354405	3354242	384	3.64E-116
wzy	-116	Intergenic	2368996..2369247	2369093	252	1.20E-320	2368912..2369293	2369124	382	5.19E-111
CE10_5175	0	Intragenic	5280860..5281131	5281034	272	6.89E-319	5280803..5281146	5280973	344	6.96E-73
lpfA	-13	Intergenic	4456513..4456771	4456674	259	6.76E-318	4456436..4456798	4456603	363	4.85E-76
wzx	0	Intergenic	2373675..2373872	2373770	198	1.03E-315	2373570..2373904	2373739	335	1.04E-74
slp	-116	Intergenic	4102360..4102572	4102475	213	4.15E-312	4101932..4102290	4102101	359	9.09E-99
envR	-40	Intergenic	3874501..3874703	3874606	203	6.27E-311	3874440..3874838	3874610	399	1.01E-73
setC	-60	Intergenic	4358138..4358358	4358235	221	4.00E-300	4358063..4358415	4358232	353	1.34E-56

sfmA	-14	Intergenic	561287..561490	561385	204	2.03E-298	561191..561547	561360	357	1.03E-65
CE10_3398	-118	Intergenic	3470862..3471063	3470966	202	4.24E-297	3470796..3471222	3470966	427	1.14E-55
CE10_4363	504	Intragenic	4445578..4445794	4445697	217	1.65E-293	4445543..4445908	4445713	366	8.52E-110
yhiM	0	Intergenic	4083276..4083495	4083398	220	4.91E-289	4083128..4083475	4083298	348	7.63E-93
ygiL	0	Intergenic	3641678..3641883	3641776	206	7.17E-286	3641600..3641975	3641769	376	2.33E-90
CE10_3627	-18	Intergenic	3707654..3707851	3707755	198	4.22E-284	3707493..3707888	3707719	396	1.23E-73
yhaI	0	Intergenic	3713998..3714219	3714096	222	1.23E-276	3713925..3714268	3714095	344	1.69E-58
eivC	-437	Intragenic	3360339..3360630	3360437	292	7.47E-275	3360325..3360830	3360495	506	3.90E-64
CE10_0067	-52	Intragenic	79557..79762	79665	206	5.62E-273	79497..79989	79666	493	3.48E-111
pitB	-45	Intergenic	3592013..3592252	3592155	240	8.54E-273	3591908..3592286	3592117	379	9.69E-67
espY4	0	Intragenic	4441633..4441849	4441731	217	1.33E-270	4439564..4439923	4439734	360	7.93E-101
yqeI	-39	Intergenic	3340559..3340755	3340657	197	2.99E-270	3340364..3340802	3340634	439	4.91E-67
torY	-12	Intergenic	2176815..2177022	2176913	208	1.36E-264	2176745..2177084	2176915	340	1.67E-57
fepE	-80	Intergenic	638504..638708	638602	205	1.92E-262	638447..638874	638617	428	2.12E-33
CE10_4312	398	Intragenic	4390373..4390601	4390505	229	7.09E-262	4390143..4390586	4390417	444	5.52E-88
yfaL	-40	Intergenic	2651310..2651511	2651408	202	7.95E-256	2651224..2651650	2651394	427	3.47E-87
prgI	-96	Intragenic	3350520..3350752	3350656	233	5.93E-251	3350469..3350830	3350638	362	5.96E-100
yhaC	716	Intragenic	3730344..3730635	3730538	292	2.88E-248	3730245..3730703	3730535	459	3.01E-110
yadN	-146	Intergenic	167088..167288	167192	201	3.31E-243	166959..167353	167184	395	1.95E-85
epaO	-41	Intergenic	3356717..3356947	3356814	231	5.27E-237	3356528..3356943	3356774	416	1.14E-79
CE10_4466	-433	Intragenic	4553360..4553551	4553453	192	5.16E-236	4553249..4553607	4553438	359	6.84E-91
yfcV	-209	Intergenic	2769906..2770124	2770029	219	5.43E-236	2769710..2770190	2770021	481	3.64E-89
speF	-416	Intergenic	730782..730985	730888	204	2.45E-235	730580..731027	730858	448	2.97E-58
rfaY	0	Intragenic	4261842..4262076	4261940	235	5.12E-234	4261768..4262126	4261938	359	5.43E-75
CE10_4598	0	Intergenic	4685949..4686161	4686047	213	8.07E-234	4685820..4686203	4686034	384	3.02E-58
garP	0	Intragenic	3735586..3735781	3735684	196	5.16E-231	3735500..3735905	3735670	406	9.92E-53
tnaA	0	Intergenic	4430400..4430614	4430498	215	9.00E-231	4430100..4430607	4430438	508	1.44E-51
yqiK	0	Intergenic	3647840..3648034	3647938	195	2.68E-229	3647709..3648091	3647922	383	9.47E-43
CE10_4228	-155	Intergenic	4305094..4305309	4305212	216	2.04E-228	4305017..4305384	4305215	368	4.03E-73

yghT	0	Intergenic	3589574..3589779	3589672	206	4.42E-228	3589494..3589850	3589681	357	5.52E-50
yiiG	0	Intergenic	4651447..4651642	4651545	196	7.59E-228	4651380..4651744	4651550	365	2.49E-47
CE10_4613	0	Intergenic	4703832..4704039	4703929	208	7.96E-223	4703697..4704049	4703880	353	9.86E-58
CE10_3508	0	Intergenic	3578743..3578967	3578870	225	1.03E-222	3578557..3579003	3578834	447	7.73E-57
ytcA	-32	Intergenic	4904288..4904497	4904384	210	4.02E-221	4904108..4904552	4904384	445	3.10E-42
kpsF	-306	Intergenic	3533457..3533707	3533614	251	2.85E-220	3533443..3533784	3533615	342	2.02E-101
cmtB	0	Intragenic	3450039..3450329	3450233	291	4.82E-218	3449910..3450401	3450233	492	6.04E-73
sipB	0	Intergenic	4362810..4363009	4362913	200	4.07E-216	4362523..4363030	4362861	508	1.28E-67
ybcK	0	Intragenic	569810..570028	569907	219	4.50E-216	569765..570178	569935	414	1.87E-67
CE10_4997	-145	Intergenic	5106630..5106836	5106727	207	1.89E-212	5106531..5106937	5106701	407	1.17E-69
yghJ	0	Intergenic	3567481..3567702	3567579	222	2.98E-212	3567439..3567947	3567609	509	6.03E-53
fdrA	-40	Intergenic	549671..549893	549796	223	4.04E-211	549602..549982	549816	381	1.17E-48
yfgH	0	Intergenic	2938103..2938308	2938211	206	3.32E-208	2938013..2938372	2938180	360	1.28E-42
espX1	0	Intergenic	28494..28696	28587	203	2.06E-208	28364..28727	28558	364	1.72E-93
ecpD	0	Intergenic	166165..166403	166306	239	4.46E-207	166136..166494	166305	359	1.29E-51
caiT	0	Intergenic	50069..50289	50192	221	5.56E-199	50350..50667	50498	318	7.46E-57
yicO	0	Intergenic	4380906..4381144	4381052	239	2.01E-198	4380817..4381160	4380987	344	3.12E-51
CE10_4567	-102	Intergenic	4657609..4657825	4657707	217	2.19E-196	4657528..4657918	4657698	391	1.82E-46
yhfL	0	Intergenic	3944903..3945145	3945001	243	4.41E-196	3944872..3945215	3945042	344	4.10E-55
yliE	796	Intragenic	908996..909193	909096	198	1.42E-195	908922..909264	909092	343	7.78E-52
yiaY	-38	Intergenic	4222519..4222737	4222640	219	1.65E-194	4222372..4222797	4222628	426	5.94E-51
fimE	-169	Intergenic	5174629..5174833	5174736	205	3.78E-194	5174540..5174936	5174710	397	1.07E-86
sat	0	Intergenic	3517517..3517729	3517614	213	1.46E-192	3517436..3517832	3517606	397	1.07E-58
agaB	0	Intergenic	3746962..3747189	3747092	228	5.33E-191	3746910..3747252	3747086	343	2.30E-37
yajR	-72	Intergenic	429382..429611	429480	230	1.44E-190	429398..429750	429581	353	2.38E-66
hutU	-51	Intergenic	769569..769824	769727	256	1.87E-190	769528..769895	769726	368	4.87E-62
neuA	0	Intragenic	3546798..3547023	3546926	226	7.23E-187	3546711..3547074	3546941	364	4.16E-62
adiY	-41	Intergenic	4937233..4937436	4937331	204	1.46E-184	4937018..4937510	4937348	493	1.84E-47
ygcG	0	Intergenic	3256315..3256517	3256420	203	3.92E-183	3256239..3256739	3256409	501	3.71E-60

fucA	-74	Intergenic	3286052..3286271	3286149	220	1.64E-182	3286026..3286389	3286220	364	5.68E-67
CE10_3076	0	Intergenic	3138278..3138518	3138421	241	1.80E-182	3138028..3138521	3138197	494	8.55E-77
kpsM	-524	Intergenic	3550543..3550703	3550600	161	3.08E-181	3550428..3550768	3550597	341	2.02E-50
CE10_0335	-219	Intergenic	370073..370290	370171	218	3.30E-180	370014..370522	370184	509	3.69E-34
CE10_0889	0	Intergenic	941040..941265	941168	226	1.17E-180	941351..941810	941515	460	9.99E-86
fucP	133	Intragenic	3286700..3286946	3286798	247	6.66E-178	3286645..3286999	3286811	355	9.10E-35
ycdT	-65	Intergenic	1166824..1167033	1166936	210	9.38E-178	1166755..1167115	1166925	361	5.86E-60
CE10_0818	0	Intragenic	865714..865932	865835	219	1.77E-176	865675..866139	865844	465	4.21E-80
ycbQ	-5	Intergenic	1039210..1039454	1039357	245	6.27E-175	1039140..1039489	1039310	350	5.18E-70
nmpC	0	Intergenic	2501916..2502198	2502014	283	1.89E-174	2501889..2502245	2502076	357	2.96E-64
csiD	-377	Intergenic	3140357..3140564	3140454	208	6.51E-174	3140288..3140657	3140458	370	5.22E-44
CE10_3557	478	Intragenic	3627437..3627729	3627535	293	6.87E-174	3627343..3627745	3627512	403	6.65E-55
fimZ	398	Intragenic	567147..567345	567245	199	1.06E-172	567095..567597	567265	503	1.49E-38
iraP	-31	Intergenic	380127..380352	380255	226	1.63E-172	380050..380418	380220	369	1.43E-58
yfbL	0	Intergenic	2697449..2697698	2697547	250	4.75E-172	2697376..2697715	2697546	340	8.05E-51
CE10_5091	-2	Intergenic	5210721..5210948	5210852	228	1.60E-171	5210688..5211031	5210858	344	1.44E-85
leuO	-181	Intergenic	92770..92982	92887	213	5.50E-171	92719..93064	92885	346	5.60E-61
gadA	-136	Intergenic	4125363..4125562	4125465	200	1.99E-167	4125231..4125617	4125450	387	1.35E-36
ibrA	-245	Intergenic	2313456..2313668	2313582	213	4.87E-166	2313396..2313798	2313629	403	1.05E-77
gadX	-69	Intergenic	4123627..4123890	4123794	264	3.56E-162	4123422..4123929	4123760	508	3.30E-51
dusA	-308	Intergenic	4862982..4863269	4863176	288	4.14E-161	4862851..4863319	4863150	469	8.07E-79
yeiT	-45	Intergenic	2539401..2539612	2539515	212	1.15E-159	2539367..2539714	2539545	348	3.52E-32
yjeJ	0	Intergenic	4993629..4993865	4993771	237	3.01E-159	4993556..4993922	4993726	367	1.55E-42
CE10_4295	0	Intergenic	4364825..4365038	4364941	214	3.42E-159	4364740..4365136	4364908	397	6.97E-59
yebN	-218	Intergenic	2124196..2124399	2124302	204	1.17E-158	2123968..2124471	2124302	504	1.15E-41
yehD	0	Intergenic	2455198..2455489	2455392	292	5.01E-157	2455067..2455497	2455329	431	1.36E-64
CE10_0523	-40	Intergenic	575784..576005	575882	222	7.25E-155	575775..576281	575943	507	6.50E-44
tdcR	0	Intragenic	3728505..3728731	3728638	227	1.96E-153	3728426..3728879	3728593	454	4.00E-55
CE10_4234	0	Intragenic	4313033..4313233	4313131	201	2.02E-151	4312961..4313314	4313131	354	2.88E-56

ysdS	0	Intergenic	2249761..2250048	2249951	288	5.19E-146	2249711..2250083	2249914	373	5.24E-63
eivF	-67	Intergenic	3365626..3365891	3365724	266	5.28E-144	3365542..3365916	3365747	375	1.33E-57
CE10_4553	0	Intergenic	4641285..4641558	4641383	274	6.98E-143	4641242..4641633	4641404	392	1.34E-62
yiaT	0	Intergenic	4215061..4215279	4215182	219	5.40E-142	4214988..4215356	4215187	369	3.42E-45
CE10_3277	0	Intragenic	3342009..3342207	3342107	199	7.34E-142	3341909..3342253	3342079	345	1.84E-61
CE10_3982	0	Intragenic	4045106..4045312	4045203	207	1.22E-141	4045036..4045387	4045219	352	6.70E-42
matA	-95	Intergenic	280175..280390	280295	216	4.62E-141	280131..280478	280301	348	4.97E-63
CE10_4519	0	Intragenic	4605362..4605587	4605490	226	1.43E-140	4605298..4605800	4605464	503	4.75E-33
ydeK	-129	Intergenic	1743172..1743407	1743311	236	2.12E-138	1743100..1743465	1743267	366	2.67E-58
eivG	0	Intragenic	3364690..3364893	3364787	204	6.94E-137	3364593..3364953	3364758	361	1.79E-38
ybcM	0	Intragenic	571111..571326	571229	216	1.92E-136	571058..571422	571253	365	1.33E-43
CE10_0264	-129	Intergenic	289056..289330	289154	275	9.27E-136	289032..289427	289202	396	5.29E-36
ybdN	-50	Intergenic	656899..657119	656997	221	9.56E-133	656777..657283	656947	507	2.34E-43
msbB	0	Intergenic	5142221..5142418	5142322	198	1.67E-132	5142143..5142485	5142313	343	5.59E-36
yegH	-88	Intergenic	2401797..2402036	2401895	240	2.68E-132	2401624..2402086	2401919	463	2.66E-45
stpA	0	Intergenic	3150441..3150636	3150539	196	2.77E-132	3150362..3150767	3150531	406	3.51E-38
CE10_2689	-7	Intergenic	2734495..2734703	2734593	209	4.65E-132	2734379..2734787	2734618	409	1.60E-46
CE10_5099	-189	Intergenic	5220663..5220891	5220794	229	8.19E-130	5220619..5221045	5220788	427	9.50E-53
ais	0	Intergenic	2677370..2677597	2677500	228	8.62E-130	2677304..2677649	2677480	346	7.77E-38
CE10_0734	0	Intergenic	789835..790057	789932	223	3.57E-126	789781..790118	789951	338	3.55E-35
CE10_0366	0	Intragenic	403466..403659	403562	194	1.01E-124	403407..403808	403575	402	2.50E-56
rpmE2	-259	Intergenic	281827..282055	281958	229	7.77E-124	281766..282121	281935	356	6.00E-44
yjcS	0	Intergenic	4906402..4906633	4906536	232	1.54E-122	4906293..4906678	4906510	386	1.42E-20
eutS	-77	Intergenic	2884428..2884644	2884526	217	1.54E-121	2884321..2884700	2884491	380	7.52E-21
ykgI	0	Intergenic	296714..296948	296852	235	5.29E-121	296520..296971	296802	452	3.14E-49
ydeP	-21	Intergenic	1733197..1733430	1733295	234	1.76E-120	1733150..1733514	1733320	365	1.09E-39
ycbR	0	Intergenic	1039926..1040129	1040024	204	4.40E-118	1039865..1040268	1040029	404	2.23E-25
CE10_1660	0	Intergenic	1691377..1691573	1691470	197	6.44E-116	1690525..1690914	1690695	390	2.80E-52
CE10_3790	-120	Intergenic	3866170..3866358	3866272	189	2.74E-116	3866095..3866520	3866265	426	2.63E-39

ugd	-66	Intergenic	2361658..2361857	2361756	200	1.90E-115	2361443..2361917	2361748	475	1.01E-34
yeeN	-21	Intergenic	2298060..2298291	2298196	232	2.40E-115	2297991..2298362	2298196	372	2.61E-48
fimB	-128	Intergenic	5173598..5173794	5173696	197	1.51E-114	5173458..5173870	5173701	413	9.51E-54
sfmC	0	Intergenic	562157..562355	562258	199	6.27E-114	562039..562383	562208	345	4.47E-24
ttdR	344	Intragenic	3660788..3661007	3660910	220	2.18E-113	3660728..3661075	3660906	348	4.99E-25
glgS	-21	Intergenic	3647068..3647359	3647262	292	5.66E-112	3647009..3647421	3647179	413	2.37E-36
CE10_3626	0	Intragenic	3706899..3707188	3706995	290	5.89E-112	3706849..3707273	3707017	425	3.39E-48
frc	-189	Intergenic	2806933..2807132	2807029	200	1.08E-111	2806858..2807355	2807027	498	3.98E-33
yegR	0	Intergenic	2432662..2432859	2432760	198	1.90E-110	2432420..2432900	2432733	481	2.75E-31
CE10_1948	132	Intergenic	1971088..1971324	1971227	237	5.92E-110	1970982..1971354	1971152	373	2.15E-60
CE10_4706	0	Intergenic	4811342..4811545	4811448	204	6.76E-108	4811275..4811649	4811440	375	6.56E-19
ybdO	0	Intergenic	657859..658149	658053	291	1.45E-107	657677..658144	657975	468	1.13E-55
rfbC	0	Intragenic	2374416..2374657	2374560	242	1.47E-106	2374347..2374712	2374543	366	1.22E-38
focA	-136	Intergenic	994945..995227	995042	283	2.04E-106	994832..995219	995053	388	2.10E-47
CE10_0066	-319	Intergenic	78407..78640	78505	234	3.75E-106	78828..79327	78997	500	1.85E-45
yebB	0	Intergenic	2165015..2165237	2165140	223	3.18E-105	2164914..2165333	2165084	420	2.72E-36
rcsA	-174	Intergenic	2239506..2239729	2239604	224	1.76E-104	2239450..2239788	2239620	339	6.76E-26
CE10_1671	0	Intragenic	1703020..1703237	1703140	218	2.06E-103	1702934..1703307	1703104	374	8.84E-29
ypdI	0	Intergenic	2808200..2808425	2808328	226	7.97E-103	2808133..2808473	2808306	341	7.17E-50
ymgG	-64	Intergenic	1355774..1356066	1355872	293	5.18E-100	1355756..1356137	1355968	382	3.68E-52
yobF	-398	Intergenic	2126920..2127135	2127038	216	1.83E-99	2126725..2127180	2127011	456	1.04E-34
CE10_4717	-163	Intragenic	4821040..4821329	4821234	290	2.00E-98	4820987..4821333	4821157	347	2.37E-27
nanC	0	Intergenic	5172449..5172718	5172547	270	2.52E-98	5172384..5172835	5172554	452	3.17E-49
hlyE	0	Intergenic	1363183..1363379	1363279	197	7.77E-98	1363069..1363423	1363254	355	3.71E-42
yfdF	0	Intragenic	2776998..2777270	2777173	273	8.77E-98	2776795..2777230	2776961	436	7.15E-51
yahL	0	Intergenic	326795..327027	326893	233	1.72E-96	326738..327153	326908	416	2.19E-42
CE10_1661	-8	Intergenic	1692752..1692955	1692849	204	2.21E-94	1692610..1692983	1692814	374	8.73E-41
nanR	-71	Intergenic	3825000..3825209	3825116	210	3.06E-93	3824959..3825338	3825169	380	2.19E-34
yjdA	-14	Intergenic	4926152..4926382	4926250	231	4.06E-93	4926072..4926458	4926239	387	3.31E-38

yiiE	-72	Intergenic	4643647..4643845	4643748	199	4.16E-93	4643595..4643982	4643765	388	2.08E-38
dpiB	-28	Intergenic	670822..671031	670934	210	3.28E-92	670615..671120	670952	506	8.14E-53
CE10_0275	0	Intragenic	303267..303554	303457	288	1.47E-91	303067..303573	303404	507	4.24E-43
feoA	-134	Intergenic	3980944..3981174	3981042	231	5.68E-91	3980903..3981321	3981073	419	1.07E-19
ybbW	0	Intergenic	540830..541052	540928	223	1.87E-90	540772..541143	540942	372	4.05E-27
CE10_0560	185	Intergenic	604888..605102	605005	215	8.20E-90	604795..605170	604965	376	5.60E-39
CE10_0290	0	Intergenic	320948..321212	321041	265	3.26E-86	320887..321231	321063	345	9.13E-39
CE10_0724	-72	Intergenic	777616..777815	777712	200	2.72E-86	777554..777900	777723	347	8.78E-51
ompN	-2	Intergenic	1605511..1605798	1605702	288	1.11E-85	1605472..1605820	1605642	349	1.44E-41
CE10_4709	0	Intergenic	4813131..4813407	4813229	277	4.69E-83	4813065..4813417	4813250	353	4.26E-21
CE10_3786	35	Intragenic	3862473..3862763	3862666	291	5.35E-82	3862492..3862833	3862662	342	3.94E-27
CE10_3301	53	Intragenic	3367459..3367708	3367557	250	1.32E-80	3366310..3366720	3366556	411	1.06E-40
yiaO	0	Intragenic	4204947..4205166	4205070	220	1.22E-77	4204828..4205239	4205070	412	1.07E-24
ydjO	0	Intergenic	2033450..2033735	2033638	286	7.78E-74	2033079..2033399	2033253	321	8.73E-58
yhbX	345	Intragenic	3778658..3778875	3778756	218	8.12E-74	3778577..3778957	3778747	381	3.79E-28
tsx2	0	Intergenic	4700453..4700681	4700551	229	5.95E-72	4700384..4700722	4700553	339	5.83E-26
yfgF	-101	Intragenic	2937553..2937750	2937653	198	7.91E-72	2937413..2937810	2937641	398	2.80E-31
araE	-58	Intergenic	3334534..3334729	3334632	196	2.18E-71	3334435..3334801	3334632	367	1.26E-21
agaC	0	Intragenic	3747549..3747788	3747647	240	3.24E-70	3747454..3747902	3747735	449	9.92E-12
cyoA	-269	Intergenic	436794..436990	436892	197	4.01E-70	436528..437034	436865	507	1.87E-25
pagP	0	Intergenic	675237..675495	675335	259	3.40E-70	675121..675484	675316	364	7.51E-49
ybeF	-190	Intergenic	680395..680594	680499	200	2.07E-69	679992..680353	680162	362	1.15E-34
yjgN	0	Intragenic	5110033..5110322	5110225	290	3.40E-68	5109929..5110378	5110099	450	1.01E-36
yahA	-152	Intergenic	310056..310291	310195	236	3.32E-65	309708..310090	309878	383	6.96E-30
ydjE	769	Intragenic	2071605..2071861	2071764	257	5.06E-64	2071601..2071970	2071771	370	1.14E-26
CE10_0258	0	Intergenic	283608..283893	283706	286	2.65E-64	283551..283948	283779	398	1.13E-42
yceJ	0	Intragenic	1190137..1190368	1190234	232	3.88E-63	1190093..1190540	1190263	448	3.09E-22
yciD	-271	Intergenic	1500835..1501071	1500974	237	3.97E-62	1500310..1500676	1500509	367	8.60E-22
mocA	0	Intragenic	3383931..3384148	3384051	218	1.57E-61	3383875..3384289	3384044	415	1.92E-09

CE10_5101	0	Intragenic	5223835..5224127	5223933	293	2.44E-59	5223824..5224191	5223994	368	1.94E-14
sfmH	159	Intragenic	565835..566060	565930	226	1.15E-55	565700..566062	565869	363	1.34E-17
gadB	-153	Intergenic	1718885..1719158	1719062	274	2.52E-54	1718763..1719137	1718933	375	3.85E-29
yffB	0	Intergenic	2899548..2899793	2899696	246	7.69E-54	2899433..2899823	2899602	391	1.12E-30
neuO	269	Intragenic	839812..840043	839946	232	5.43E-50	839606..839902	839776	297	1.89E-19
CE10_4222	0	Intergenic	4299049..4299300	4299203	252	8.13E-48	4298934..4299442	4299104	509	2.55E-18

Table S-5 CE10 DsdC2 binding sites (no D-ser).

			Replicate 1				Replicate 2			
Gene	**Dist. Gene**	**Context**	**Co-ordinates**	**Centre**	**Length**	***p*-value**	**Co-ordinates**	**Centre**	**Length**	***p*-value**
CE10_0023	-534	Intragenic	27362..27811	27681	450	0.00E+00	27442..27871	27704	430	9.35E-213
ycgX	-686	Intergenic	1342133..1342499	1342311	367	0.00E+00	1342023..1342527	1342193	505	1.23E-214
wbbC	0	Intragenic	2367810..2368203	2367985	394	0.00E+00	2367814..2368253	2367984	440	0.00E+00
dsdC1	0	Intergenic	2791240..2791618	2791420	379	0.00E+00	2791226..2791600	2791395	375	0.00E+00
waaT	0	Intragenic	4262774..4263197	4263021	424	0.00E+00	4262768..4263185	4263017	418	8.06E-166
dsdC2	0	Intergenic	4332279..4332649	4332458	371	0.00E+00	4332276..4332631	4332446	356	0.00E+00
CE10_4270	0	Intragenic	4347265..4347784	4347430	520	0.00E+00	4347264..4347758	4347434	495	7.32E-239
CE10_4363	1061	Intragenic	4444704..4445178	4445027	475	0.00E+00	4445502..4445873	4445654	372	1.25E-237
lpfA	0	Intergenic	4456422..4456804	4456599	383	0.00E+00	4456392..4456771	4456628	380	1.11E-108
vioA	0	Intragenic	2372516..2372867	2372689	352	8.90E-323	2372423..2372664	2372593	242	1.76E-150
wzy	0	Intergenic	2368896..2369216	2369049	321	1.82E-315	2368871..2369265	2369173	395	4.50E-193
neuB	-421	Intergenic	3548330..3548635	3548441	306	1.37E-302	3548251..3548595	3548412	345	9.22E-194
CE10_2740	-18	Intergenic	2788678..2789063	2788885	386	4.43E-302	2788665..2789166	2789003	502	7.81E-172
waaV	0	Intergenic	4259991..4260437	4260258	447	7.38E-296	4260082..4260441	4260280	360	2.76E-188

rfaL	499	Intragenic	4258236..4258626	4258447	391	2.78E-295	4258276..4258652	4258442	377	1.61E-154
epaS	188	Intragenic	3354065..3354430	3354251	366	5.54E-287	3354022..3354464	3354304	443	8.69E-206
ubiC	706	Intragenic	4853339..4853690	4853591	352	1.30E-281	4853266..4853700	4853584	435	7.05E-211
yfcV	-55	Intergenic	2769752..2770201	2769931	450	1.33E-278	2769693..2770190	2770021	498	1.69E-116
yjbM	0	Intragenic	4862245..4862634	4862363	390	7.55E-274	4862245..4862639	4862414	395	1.45E-131
yjiC	0	Intergenic	5189271..5189805	5189628	535	1.45E-264	5189302..5189756	5189464	455	1.38E-162
espX1	0	Intergenic	28370..28773	28600	404	1.10E-262	28209..28538	28287	330	6.41E-134
ygiL	0	Intergenic	3641569..3642041	3641733	473	3.65E-259	3641572..3642003	3641740	432	1.39E-111
yhaC	667	Intragenic	3730287..3730712	3730466	426	4.99E-252	3730447..3730759	3730597	313	2.64E-165
CE10_4160	0	Intragenic	4233464..4233945	4233766	482	2.26E-250	4233513..4233854	4233685	342	1.41E-179
ybjQ	-414	Intragenic	942239..942775	942418	537	2.53E-243	942545..942818	942645	274	5.34E-117
CE10_0067	0	Intergenic	79472..79809	79662	338	4.14E-243	79624..80018	79862	395	7.90E-167
tauA	-100	Intergenic	365174..365547	365353	374	1.34E-241	365200..365553	365384	354	1.45E-127
espX4	0	Intragenic	4852588..4853001	4852768	414	4.71E-239	4852613..4852982	4852813	370	1.46E-150
yjbE	0	Intergenic	4839307..4839675	4839496	369	6.74E-237	4839305..4839662	4839494	358	1.31E-110
hutU	-14	Intergenic	769437..769861	769682	425	1.17E-236	769481..769849	769651	369	2.76E-105
CE10_4870	0	Intergenic	4974849..4975240	4975061	392	5.12E-234	4974865..4975238	4975086	374	2.38E-104
yhhZ	0	Intergenic	4028972..4029264	4029151	293	3.40E-231	4029028..4029274	4029198	247	5.48E-104
CE10_4297	0	Intergenic	4367341..4367731	4367502	391	2.40E-228	4367294..4367769	4367619	476	7.47E-108
kpsF	-102	Intergenic	3533379..3533911	3533557	533	3.43E-227	3533361..3533868	3533531	508	6.94E-156
ibrA	-228	Intergenic	2313439..2313870	2313580	432	2.33E-226	2313352..2313794	2313522	443	2.46E-114
dusA	-227	Intragenic	4862943..4863300	4863027	358	6.78E-226	4862921..4863413	4863244	493	8.59E-143
CE10_4943	0	Intergenic	5047548..5047940	5047770	393	6.65E-225	5047602..5047974	5047772	373	4.37E-143
CE10_0818	0	Intragenic	865658..866109	865838	452	9.36E-203	865680..866090	865850	411	5.01E-107
yqeI	0	Intergenic	3340261..3340795	3340619	535	5.84E-201	3340376..3340781	3340630	406	2.98E-106
sipB	0	Intergenic	4362531..4363052	4362886	522	1.32E-200	4362545..4362997	4362828	453	1.78E-113
yfaL	0	Intergenic	2651212..2651655	2651388	444	3.05E-198	2650598..2651022	2650754	425	2.54E-62

cmtB	0	Intragenic	3449979..3450401	3450155	423	3.05E-193	3449934..3450266	3450154	333	6.45E-75
leuO	-103	Intergenic	92697..93060	92877	364	2.90E-190	92685..93057	92888	373	2.35E-108
sfmA	0	Intergenic	561144..561543	561365	400	9.68E-188	561186..561544	561375	359	5.48E-105
nmpC	0	Intergenic	2501882..2502248	2502062	367	9.00E-185	2501915..2502252	2502082	338	4.42E-84
envR	0	Intergenic	3874432..3874842	3874602	411	3.50E-184	3874441..3874892	3874608	452	1.39E-107
CE10_1948	0	Intergenic	1970999..1971344	1971161	346	6.71E-179	1970975..1971377	1971145	403	9.16E-88
ykgI	0	Intragenic	296587..297000	296821	414	1.03E-177	296534..296957	296802	424	5.01E-114
CE10_4228	-48	Intergenic	4305007..4305416	4305187	410	2.78E-177	4305007..4305370	4305176	364	3.90E-117
yadN	-20	Intergenic	166962..167347	167171	386	2.17E-175	166987..167325	167156	339	2.42E-149
yhaI	0	Intergenic	3713935..3714331	3714114	397	5.08E-174	3713932..3714346	3714096	415	7.53E-92
CE10_2939	0	Intergenic	2996428..2996801	2996622	374	1.64E-172	2996395..2996787	2996556	393	1.02E-72
ecpD	0	Intergenic	166089..166455	166319	367	5.10E-171	166137..166580	166307	444	1.51E-69
ysdS	-2	Intergenic	2249696..2249982	2249774	287	9.91E-170	2249697..2250095	2249928	399	8.53E-103
fimE	0	Intergenic	5174540..5175005	5174719	466	2.90E-167	5174553..5174991	5174706	439	1.24E-115
eivF	0	Intergenic	3365537..3365942	3365766	406	3.43E-167	3365476..3365951	3365641	476	2.06E-86
CE10_5099	-53	Intergenic	5220615..5221027	5220787	413	4.82E-167	5220846..5221065	5220944	220	3.91E-60
CE10_2689	0	Intergenic	2734378..2734801	2734629	424	8.09E-166	2734336..2734811	2734648	476	1.47E-88
CE10_4295	0	Intragenic	4364621..4365097	4364979	477	8.67E-166	4364636..4365088	4364974	453	7.28E-86
yfdF	0	Intergenic	2776752..2777229	2776930	478	2.97E-164	2776797..2777268	2776946	472	2.01E-82
ycbQ	0	Intergenic	1039076..1039514	1039335	439	5.13E-162	1039143..1039484	1039313	342	7.36E-104
yiiG	0	Intergenic	4651354..4651747	4651569	394	9.84E-162	4651373..4651738	4651542	366	6.90E-60
CE10_3508	0	Intergenic	3578622..3579000	3578826	379	6.22E-154	3578596..3578958	3578866	363	4.72E-77
CE10_3627	0	Intergenic	3707535..3707897	3707711	363	1.21E-152	3707473..3707903	3707634	431	8.84E-101
ymgG	0	Intergenic	1355705..1356177	1355885	473	3.86E-152	1355703..1356174	1355872	472	1.75E-83
fucA	-48	Intergenic	3286026..3286411	3286206	386	8.73E-149	3285981..3286374	3286151	394	2.09E-114
torY	0	Intergenic	2176727..2177084	2176887	358	2.30E-148	2176744..2177192	2176914	449	6.53E-80
kpsM	-372	Intergenic	3550391..3550760	3550608	370	2.32E-148	3550432..3550730	3550509	299	1.12E-112

matA	-57	Intergenic	280137..280561	280314	425	7.16E-147	280121..280598	280262	478	2.11E-86
CE10_4598	0	Intragenic	4685849..4686220	4686045	372	1.38E-146	4685779..4686266	4686099	488	2.64E-60
ydjO	225	Intragenic	2033083..2033319	2033244	237	4.06E-146	2033342..2033737	2033576	396	3.47E-90
yfbL	0	Intergenic	2697381..2697743	2697560	363	2.99E-145	2697379..2697799	2697547	421	1.31E-74
ygcG	0	Intergenic	3256240..3256635	3256419	396	6.84E-145	3256265..3256590	3256454	326	2.13E-89
yhfL	0	Intergenic	3944842..3945211	3945021	370	2.46E-144	3944850..3945263	3945015	414	2.95E-107
ycdT	0	Intergenic	1166750..1167135	1166927	386	2.46E-142	1166695..1167145	1166976	451	2.84E-91
yliE	650	Intragenic	908860..909306	909139	447	1.70E-139	908896..909299	909149	404	2.95E-70
CE10_4613	0	Intergenic	4703706..4704134	4703859	429	1.40E-137	4703716..4704119	4703886	404	1.55E-87
ydeK	-14	Intergenic	1743057..1743518	1743342	462	2.95E-136	1743081..1743453	1743251	373	1.87E-63
yiaY	0	Intergenic	4222409..4222788	4222589	380	3.93E-136	4222384..4222793	4222553	410	8.17E-72
caiT	0	Intergenic	50024..50329	50201	306	4.09E-135	50023..50271	50191	249	7.11E-115
iraP	0	Intergenic	380060..380434	380239	375	5.10E-135	379991..380420	380286	430	2.05E-79
yajR	-8	Intergenic	429318..429824	429492	507	9.68E-135	429366..429789	429533	424	4.88E-111
yghJ	0	Intragenic	3567434..3567898	3567611	465	2.95E-134	3567470..3567875	3567638	406	1.60E-103
pagP	-249	Intergenic	674770..675139	674944	370	3.87E-132	675168..675524	675334	357	2.58E-65
CE10_4234	0	Intragenic	4312928..4313380	4313105	453	2.38E-131	4312961..4313301	4313131	341	3.97E-79
pitB	0	Intergenic	3591899..3592395	3592078	497	1.32E-129	3591909..3592352	3592183	444	4.22E-66
ybdO	0	Intergenic	657734..658132	657967	399	2.06E-129	657795..658194	658025	400	2.99E-64
rfbC	0	Intragenic	2374333..2374738	2374559	406	2.65E-128	2374329..2374695	2374569	367	1.62E-59
setC	0	Intergenic	4358051..4358447	4358274	397	1.32E-127	4358061..4358437	4358231	377	7.13E-79
ais	0	Intergenic	2677273..2677731	2677556	459	1.78E-126	2677300..2677681	2677521	382	1.81E-38
CE10_3398	0	Intergenic	3470812..3471198	3471026	387	6.90E-125	3470820..3471244	3470982	425	1.41E-81
hipB	-500	Intragenic	1739028..1739535	1739358	508	8.72E-125	1739163..1739521	1739352	359	2.40E-56
CE10_3626	0	Intergenic	3706823..3707193	3707003	371	2.74E-124	3706855..3707229	3707025	375	5.77E-93
fimB	0	Intergenic	5173476..5174004	5173646	529	1.21E-119	5173463..5173786	5173685	324	4.73E-47
fdrA	0	Intergenic	549610..549982	549790	373	4.95E-118	549513..549940	549801	428	2.69E-74

sat	0	Intergenic	3517461..3517916	3517634	456	1.00E-118	3517411..3517844	3517677	434	5.79E-75
tnaA	0	Intergenic	4430054..4430587	4430234	534	1.80E-117	4430209..4430597	4430434	389	1.23E-59
yqiG	0	Intragenic	3642402..3642914	3642582	513	6.12E-117	3642397..3642702	3642630	306	5.57E-80
garP	0	Intragenic	3735483..3735991	3735662	509	9.76E-116	3735486..3735929	3735640	444	3.60E-96
yghT	0	Intergenic	3589485..3589864	3589685	380	1.17E-114	3589507..3589892	3589723	386	1.65E-38
CE10_4567	0	Intergenic	4657542..4657936	4657770	395	4.47E-114	4657524..4657832	4657743	309	9.52E-70
CE10_2267	-177	Intergenic	2290168..2290519	2290432	352	2.20E-113	2290164..2290552	2290331	389	2.86E-94
CE10_3557	500	Intragenic	3627487..3627733	3627536	247	9.24E-113	3627485..3627894	3627655	410	4.27E-74
yiaT	0	Intergenic	4214955..4215397	4215127	443	3.12E-111	4214986..4215330	4215154	345	4.85E-83
yicO	0	Intergenic	4380807..4381176	4380977	370	2.84E-110	4380736..4381220	4381051	485	3.75E-65
yegH	-48	Intergenic	2401698..2402076	2401876	379	5.27E-110	2401645..2402032	2401772	388	7.60E-91
yeeN	0	Intergenic	2298026..2298400	2298206	375	7.14E-110	2297949..2298402	2298256	454	1.18E-74
CE10_0275	0	Intragenic	303207..303549	303328	343	1.07E-107	303163..303556	303388	394	9.07E-54
speF	-275	Intergenic	730641..731053	730874	413	1.91E-107	730585..731020	730853	436	9.40E-62
fimA	-167	Intergenic	5175542..5175912	5175798	371	2.15E-107	5175600..5175850	5175782	251	6.78E-76
CE10_3982	0	Intragenic	4045022..4045392	4045213	371	5.01E-107	4044997..4045435	4045146	439	2.07E-64
yfgH	0	Intergenic	2938010..2938381	2938208	372	1.47E-104	2938055..2938475	2938221	421	2.75E-64
nanC	0	Intergenic	5172434..5172888	5172605	455	1.06E-102	5172427..5172782	5172614	356	1.34E-50
CE10_1661	0	Intergenic	1692553..1693004	1692835	452	1.53E-101	1692581..1693019	1692850	439	1.74E-54
yiaW	0	Intergenic	4217591..4217912	4217775	322	9.98E-101	4217635..4218015	4217800	381	1.77E-61
dpiB	0	Intergenic	670720..671107	670863	388	1.83E-99	670710..671140	670973	431	2.79E-52
gadX	0	Intergenic	4123396..4123905	4123732	510	1.90E-99	4123414..4123921	4123752	508	1.63E-117
adiY	0	Intergenic	4937134..4937553	4937314	420	7.88E-99	4937053..4937508	4937362	456	2.93E-73
gadA	0	Intergenic	4125210..4125622	4125383	413	8.81E-99	4125220..4125686	4125373	467	2.15E-61
stpA	0	Intergenic	3150326..3150740	3150505	415	1.51E-98	3150388..3150759	3150596	372	2.76E-67
yebB	0	Intergenic	2164887..2165369	2165056	483	2.07E-98	2164900..2165404	2165070	505	1.06E-61
fucP	75	Intragenic	3286606..3287099	3286748	494	1.15E-97	3286596..3287010	3286764	415	5.54E-55

ykgE	-86	Intergenic	299401..299767	299579	367	3.46E-97	299337..299757	299588	421	1.86E-35
CE10_0735	-119	Intergenic	790984..791457	791318	474	2.25E-94	791118..791462	791299	345	3.46E-51
rpmE2	-218	Intergenic	281786..282149	281959	364	1.37E-92	281769..282123	281954	355	3.63E-47
CE10_0366	0	Intragenic	403367..403755	403545	389	4.75E-92	403415..403787	403585	373	1.88E-60
fimZ	285	Intragenic	567051..567472	567230	422	1.56E-91	567094..567464	567264	371	4.27E-76
csiD	-967	Intragenic	3139530..3139974	3139705	445	4.09E-89	3139562..3139981	3139812	420	4.07E-53
ypdI	0	Intergenic	2808118..2808499	2808324	382	3.40E-88	2808126..2808526	2808357	401	1.18E-50
yobF	-268	Intergenic	2126790..2127203	2127024	414	7.30E-88	2126793..2127148	2126962	356	2.65E-58
CE10_0724	0	Intergenic	777533..777898	777683	366	3.45E-83	777556..777950	777645	395	1.95E-45
glgS	0	Intergenic	3647029..3647437	3647258	409	6.00E-83	3647008..3647465	3647178	458	1.05E-45
yjdA	0	Intergenic	4926058..4926477	4926298	420	2.80E-82	4926030..4926402	4926289	373	3.04E-47
frc	0	Intergenic	2806899..2807342	2807163	444	1.45E-81	2806880..2807282	2807017	403	7.72E-45
yebN	-138	Intergenic	2124095..2124479	2124275	385	1.53E-81	2123997..2124498	2124330	502	7.66E-57
ytcA	0	Intergenic	4904194..4904592	4904372	399	7.52E-81	4904214..4904550	4904382	337	2.48E-61
lfpB	0	Intragenic	4455596..4455976	4455809	381	1.06E-80	4455579..4456010	4455842	432	4.25E-63
yihN	0	Intergenic	4624721..4625143	4624988	423	3.42E-78	4624732..4625142	4624976	411	7.21E-46
neuO	180	Intragenic	839731..840037	839842	307	2.50E-74	839122..839433	839361	312	2.24E-22
CE10_4222	0	Intergenic	4299056..4299433	4299236	378	1.43E-47	4299050..4299330	4299238	281	2.40E-32

Table S-6 CE10 DsdC2 binding sites (+ D-ser).

			Replicate 1				Replicate 2			
Gene	**Dist. Gene**	**Context**	**Co-ordinates**	**Centre**	**Length**	***p*-value**	**Co-ordinates**	**Centre**	**Length**	***p*-value**
dsdC1	0	Intergenic	2791201..2791623	2791424	423	0.00E+00	2791239..2791631	2791405	393	0.00E+00

dsdC2	0	Intergenic	4332245..4332661	4332462	417	0.00E+00	4332276..4332653	4332446	378	0.00E+00
wbbC	0	Intragenic	2367801..2368218	2368019	418	7.61E-246	2367849..2368187	2368018	339	0.00E+00
CE10_4270	0	Intragenic	4347302..4347887	4347499	586	1.37E-196	4347239..4347682	4347409	444	0.00E+00
CE10_4160	0	Intragenic	4233415..4233971	4233772	557	6.72E-166	4233478..4233946	4233779	469	0.00E+00
wzy	-61	Intergenic	2368941..2369336	2369143	396	9.88E-163	2368949..2369329	2369075	381	0.00E+00
CE10_0023	-567	Intragenic	27395..27867	27682	473	1.75E-161	27398..27802	27567	405	0.00E+00
neuB	-351	Intergenic	3548260..3548659	3548505	400	5.44E-160	3548210..3548563	3548377	354	0.00E+00
CE10_2740	-31	Intergenic	2788691..2789078	2788926	388	2.70E-158	2789250..2789602	2789408	353	0.00E+00
waaV	0	Intergenic	4259986..4260392	4260149	407	9.89E-158	4260036..4260393	4260224	358	0.00E+00
ygiL	0	Intergenic	3641588..3642068	3641783	481	3.73E-154	3641569..3642072	3641737	504	0.00E+00
espX4	0	Intragenic	4852560..4852976	4852760	417	1.24E-150	4852574..4852929	4852769	356	0.00E+00
rfaL	419	Intragenic	4258187..4258738	4258380	552	4.45E-146	4258266..4258601	4258436	336	0.00E+00
epaS	154	Intragenic	3353928..3354450	3354126	523	1.66E-144	3354034..3354408	3354242	375	0.00E+00
yjiC	0	Intergenic	5189241..5189815	5189623	575	2.49E-144	5189297..5189760	5189591	464	0.00E+00
ycgX	600	Intergenic	1342145..1342585	1342410	441	5.17E-144	1342107..1342478	1342309	372	0.00E+00
waaT	0	Intragenic	4262782..4263209	4263010	428	5.22E-142	4262767..4263185	4263019	419	0.00E+00
espY4	349	Intragenic	4441008..4441461	4441273	454	4.53E-140	4440975..4441325	4441190	351	2.83E-300
CE10_4870	0	Intergenic	4974839..4975298	4975038	460	9.58E-137	4974862..4975255	4975032	394	0.00E+00
CE10_0067	0	Intragenic	79527..80036	79718	510	4.12E-135	79505..79780	79674	276	0.00E+00
prgI	-15	Intragenic	3350439..3350870	3350639	432	6.54E-130	3350441..3350883	3350611	443	0.00E+00
yhhZ	0	Intergenic	4028998..4029273	4029197	276	2.42E-124	4028968..4029295	4029138	328	0.00E+00
yfaL	0	Intergenic	2651189..2651693	2651494	505	5.57E-123	2651235..2651598	2651405	364	1.92E-267
yjbM	70	Intragenic	4862267..4862581	4862385	315	3.24E-112	4862932..4863330	4863089	399	2.32E-275
kpsF	-135	Intergenic	3533449..3533878	3533679	430	8.64E-112	3533422..3533766	3533614	345	1.14E-298
ubiC	-640	Intragenic	4853358..4853756	4853561	399	8.80E-112	4853286..4853763	4853597	478	0.00E+00
CE10_4297	0	Intergenic	4367262..4367846	4367652	585	2.91E-108	4367300..4367779	4367611	480	0.00E+00
tauA	-124	Intergenic	365045..365523	365326	479	4.47E-107	365186..365528	365359	343	0.00E+00

ibrA	-150	Intergenic	2313361..2313946	2313549	586	4.04E-102	2313418..2313818	2313583	401	1.44E-293
yhaC	601	Intragenic	3730248..3730669	3730558	422	1.23E-101	3730146..3730645	3730316	500	0.00E+00
eivC	-386	Intragenic	3360288..3360700	3360502	413	7.30E-99	3360273..3360679	3360443	407	1.52E-292
yghJ	0	Intragenic	3567495..3567903	3567694	409	1.32E-96	3567412..3567851	3567579	440	5.16E-274
eprH	0	Intergenic	3351518..3352032	3351839	515	1.81E-96	3351593..3351969	3351800	377	0.00E+00
envR	0	Intergenic	3874394..3874832	3874633	439	1.33E-94	3874434..3874804	3874604	371	0.00E+00
dusA	-251	Intragenic	4862922..4863326	4863120	405	1.79E-90	4862932..4863330	4863089	399	2.32E-275
ybcK	0	Intragenic	569700..570111	569916	412	9.24E-90	569742..570133	569908	392	1.21E-293
CE10_3508	0	Intergenic	3578606..3579015	3578816	410	2.43E-87	3578606..3579032	3578863	427	2.23E-263
setC	0	Intergenic	4357955..4358444	4358246	490	8.58E-87	4358028..4358401	4358232	374	2.00E-219
espX1	0	Intergenic	28309..28800	28603	492	1.97E-87	28392..28756	28559	365	0.00E+00
sfmA	0	Intergenic	561132..561550	561351	419	1.71E-85	561191..561560	561361	370	7.04E-234
CE10_0818	-110	Intragenic	865607..866128	865806	522	3.95E-85	865643..866088	865813	446	0.00E+00
CE10_3398	-4	Intergenic	3470772..3471177	3470964	406	6.67E-84	3470777..3471175	3470947	399	7.28E-225
yjbE	0	Intergenic	4839264..4839724	4839462	461	9.84E-83	4839262..4839689	4839521	428	0.00E+00
CE10_4943	0	Intergenic	5047521..5048012	5047717	492	7.13E-83	5047518..5047942	5047773	425	0.00E+00
sipB	0	Intergenic	4362518..4363039	4362894	522	3.76E-82	4362542..4363020	4362854	479	0.00E+00
gadX	0	Intergenic	4123530..4123934	4123735	405	1.66E-81	4123408..4123914	4123745	507	1.24E-201
eivF	0	Intergenic	3365516..3365886	3365686	371	3.69E-80	3365504..3365952	3365783	449	5.75E-162
ysdS	0	Intergenic	2249552..2250108	2249911	557	4.69E-80	2249708..2250086	2249917	379	4.66E-197
CE10_4228	0	Intergenic	4304996..4305491	4305195	496	8.81E-79	4305009..4305514	4305179	506	1.88E-286
yehD	0	Intergenic	2455112..2455543	2455344	432	6.98E-78	2455141..2455485	2455307	345	2.52E-134
CE10_3557	0	Intergenic	3628241..3628807	3628437	567	2.16E-78	3628255..3628751	3628424	497	1.25E-246
rfaY	0	Intragenic	4261738..4262158	4261857	421	3.44E-75	4261801..4262123	4261980	323	1.47E-278
lpfA	0	Intergenic	4456369..4456823	4456624	455	5.57E-75	4456471..4456810	4456642	340	0.00E+00
fimZ	271	Intragenic	567067..567513	567317	447	2.79E-74	567021..567429	567266	409	1.58E-149
ycbQ	0	Intergenic	1039119..1039559	1039363	441	8.68E-74	1039098..1039524	1039356	427	2.36E-311

yiaY	0	Intergenic	4222437..4222812	4222611	376	3.07E-73	4222380..4222772	4222603	393	3.22E-220
neuA	0	Intragenic	3546713..3547161	3546913	449	4.29E-73	3546735..3547143	3546897	409	3.46E-227
cmtB	0	Intragenic	3449964..3450395	3450164	432	6.85E-73	3449903..3450404	3450235	502	4.11E-218
tnaA	0	Intergenic	4430163..4430581	4430363	419	9.87E-73	4430328..4430607	4430441	280	2.34E-184
yliE	583	Intragenic	908813..909295	909096	483	2.28E-71	908913..909264	909096	352	4.15E-157
hutU	-18	Intergenic	769262..769857	769658	596	1.78E-71	769533..769887	769718	355	0.00E+00
nmpC	0	Intergenic	2501805..2502289	2502090	485	2.12E-70	2501843..2502269	2502013	427	9.50E-291
yajR	0	Intergenic	429282..429793	429477	512	9.81E-70	429395..429749	429580	355	5.17E-257
CE10_4598	0	Intragenic	4685781..4686273	4686074	493	2.38E-70	4685806..4686215	4686046	410	9.96E-268
yadN	0	Intergenic	166932..167338	167148	407	1.12E-69	166938..167356	167187	419	0.00E+00
fimE	-4	Intergenic	5174588..5174998	5174788	411	1.28E-69	5174557..5174977	5174727	421	2.97E-284
fucA	0	Intragenic	3285938..3286459	3286138	522	2.45E-68	3285970..3286380	3286213	411	2.94E-236
CE10_4567	0	Intergenic	4657513..4657954	4657713	442	8.36E-68	4657572..4657925	4657738	354	7.83E-180
CE10_2939	0	Intergenic	2996379..2996812	2996579	434	1.72E-67	2996461..2996801	2996632	341	0.00E+00
sipD	0	Intergenic	4364806..4365170	4364944	365	2.48E-66	4365735..4366090	4365923	356	0.00E+00
yhaI	0	Intergenic	3713855..3714452	3714055	598	1.68E-65	3713910..3714266	3714080	357	0.00E+00
garP	0	Intragenic	3735491..3736089	3735691	599	1.79E-65	3735521..3735844	3735691	324	2.52E-287
CE10_2689	0	Intergenic	2734327..2734783	2734587	457	3.07E-65	2734299..2734791	2734626	493	1.40E-233
pitB	0	Intergenic	3591846..3592320	3592045	475	3.33E-64	3591924..3592289	3592120	366	3.87E-315
caiT	0	Intergenic	50113..50672	50478	560	6.57E-64	50362..50693	50524	332	4.33E-240
yegH	-1	Intergenic	2401649..2402123	2401924	475	3.00E-63	2401704..2402035	2401865	332	2.50E-178
CE10_3627	0	Intergenic	3707414..3707887	3707688	474	4.51E-63	3707478..3707905	3707736	428	5.20E-295
kpsM	-250	Intergenic	3550245..3550773	3550588	529	4.71E-62	3550359..3550819	3550650	461	3.54E-182
CE10_4613	0	Intergenic	4703675..4704137	4703869	463	8.55E-61	4703713..4704075	4703906	363	8.90E-220
sat	0	Intergenic	3517434..3517910	3517626	477	5.64E-60	3517389..3517834	3517674	446	3.14E-183
ydeP	0	Intergenic	1733125..1733538	1733339	414	1.65E-59	1733128..1733543	1733374	416	1.02E-100
focA	-42	Intergenic	994851..995268	995072	418	5.48E-58	994837..995244	995079	408	7.45E-156

ecpD	0	Intergenic	166100..166433	166301	334	7.55E-58	166101..166485	166271	385	4.73E-279
neuO	460	Intragenic	839980..840574	840375	595	2.92E-57	839682..840049	839850	368	1.84E-62
fimB	0	Intergenic	5173383..5173942	5173582	560	2.79E-55	5172313..5172781	5172613	469	9.36E-174
matA	-7	Intergenic	280087..280499	280309	413	5.70E-55	280132..280487	280302	356	2.54E-128
stpA	0	Intragenic	3150391..3150799	3150588	409	1.20E-53	3150389..3150796	3150557	408	8.81E-172
leuO	0	Intergenic	92657..93240	92857	584	1.38E-53	92696..93128	92864	433	1.97E-268
ykgE	-118	Intergenic	299238..299735	299431	498	1.88E-53	299344..299739	299582	396	1.72E-142
ycdT	0	Intergenic	1166780..1167160	1166947	381	6.48E-53	1166719..1167223	1167057	505	4.78E-190
ydeK	-11	Intergenic	1743054..1743494	1743254	441	2.64E-52	1743103..1743473	1743304	371	3.33E-245
yiaW	0	Intergenic	4217559..4218001	4217802	443	5.78E-52	4217589..4217942	4217759	354	9.29E-199
ykgI	0	Intergenic	296616..296991	296780	376	5.62E-52	296611..296965	296801	355	1.20E-119
ymgG	0	Intergenic	1355685..1356124	1355885	440	5.64E-51	1355700..1356166	1355868	467	1.60E-203
yfcV	-6	Intergenic	2769704..2770197	2769884	494	6.00E-50	2769666..2770169	2770000	504	0.00E+00
CE10_4234	0	Intragenic	4312886..4313336	4313141	451	1.57E-49	4312952..4313319	4313117	368	7.90E-259
CE10_0724	0	Intergenic	777460..777936	777740	477	1.10E-49	777542..778046	777712	505	1.50E-89
CE10_3790	0	Intergenic	3866043..3866625	3866432	583	1.02E-47	3866051..3866515	3866177	465	2.54E-94
ybcM	62	Intragenic	571029..571426	571230	398	1.17E-47	571034..571492	571201	459	6.70E-133
pagP	0	Intergenic	675200..675525	675326	326	1.18E-47	675148..675511	675275	364	4.33E-165
glgS	0	Intergenic	3646974..3647421	3647225	448	3.61E-47	3646900..3647400	3647231	501	1.02E-146
CE10_0366	0	Intragenic	403402..404000	403602	599	7.37E-47	403365..403834	403525	470	9.80E-149
yiiG	0	Intergenic	4651287..4651811	4651615	525	1.66E-46	4651381..4651749	4651582	369	1.11E-243
nanR	0	Intergenic	3824875..3825370	3825171	496	1.79E-46	3824940..3825262	3825109	323	1.35E-97
CE10_0258	0	Intergenic	283534..283987	283733	454	2.03E-46	283567..283945	283776	379	7.20E-88
yfbL	0	Intergenic	2697340..2697935	2697537	596	3.90E-46	2697375..2697816	2697545	442	1.20E-250
ttdR	270	Intragenic	3660664..3661092	3660915	429	1.03E-45	3660735..3661078	3660909	344	3.99E-113
yeaI	0	Intragenic	2088989..2089430	2089240	442	2.68E-45	2089041..2089446	2089211	406	1.11E-122
ygcG	0	Intergenic	3256217..3256779	3256399	563	2.45E-45	3256230..3256703	3256400	474	5.06E-206

yeeN	0	Intergenic	2297945..2298387	2298145	443	2.37E-44	2297960..2298355	2298198	396	7.72E-118
adiY	0	Intergenic	4937037..4937578	4937380	542	2.98E-44	4937059..4937501	4937332	443	5.50E-216
agaB	0	Intergenic	3746867..3747317	3747125	451	3.16E-44	3746875..3747245	3747078	371	5.87E-178
nanC	0	Intergenic	5172352..5172793	5172545	442	3.35E-44	5172313..5172781	5172613	469	9.36E-174
ybbW	0	Intergenic	540753..541206	540951	454	5.86E-44	540741..541217	540907	477	3.44E-86
yjfI	0	Intragenic	5028723..5029121	5028920	399	6.16E-44	5028685..5029152	5028839	468	2.05E-168
ytcA	0	Intergenic	4904147..4904575	4904347	429	1.22E-43	4904104..4904563	4904397	460	5.28E-175
yhfL	0	Intergenic	3944807..3945277	3945083	471	2.46E-43	3944846..3945244	3945014	399	0.00E+00
yiaT	0	Intergenic	4214953..4215409	4215149	457	5.93E-43	4214983..4215330	4215162	348	3.70E-220
CE10_5099	0	Intergenic	5220777..5221310	5221111	534	6.42E-43	5220623..5221088	5220791	466	1.36E-202
CE10_3982	0	Intragenic	4044927..4045465	4045270	539	9.27E-43	4045050..4045404	4045220	355	3.81E-142
iraP	0	Intergenic	379970..380442	380156	473	3.42E-42	380017..380409	380184	393	2.64E-141
ais	0	Intergenic	2677253..2677737	2677543	485	1.24E-40	2677245..2677642	2677476	398	2.37E-188
speF	-254	Intergenic	730620..731105	730906	486	1.76E-40	730699..731027	730863	329	1.83E-211
fdrA	0	Intergenic	549578..549979	549774	402	1.83E-40	549626..549978	549809	353	3.33E-180
torY	0	Intergenic	2176663..2177104	2176905	442	1.42E-39	2176728..2177090	2176896	363	9.43E-313
rpmE2	-185	Intergenic	281753..282201	282006	449	1.43E-39	281742..282113	281907	372	6.57E-194
yeiT	0	Intragenic	2539229..2539663	2539426	435	5.14E-39	2539357..2539676	2539509	320	2.29E-129
CE10_0735	-142	Intergenic	791007..791403	791214	397	5.41E-39	791122..791460	791290	339	2.71E-128
yiiE	0	Intergenic	4643638..4644071	4643838	434	2.57E-38	4643550..4644028	4643718	479	1.65E-91
gadA	0	Intergenic	4125236..4125634	4125437	399	4.55E-38	4125267..4125633	4125437	367	3.83E-176
CE10_0066	0	Intergenic	78824..79258	79024	435	7.11E-38	78777..79156	78987	380	3.96E-150
fimA	-83	Intergenic	5175568..5175996	5175763	429	9.04E-38	5175462..5175893	5175735	432	4.64E-137
CE10_1661	0	Intergenic	1692466..1693013	1692815	548	1.40E-37	1692603..1692999	1692836	397	4.86E-142
feoA	-23	Intergenic	3980886..3981285	3981088	400	3.87E-37	3980897..3981257	3981092	361	1.21E-138
siiCA	0	Intergenic	498197..498591	498397	395	3.90E-37	498178..498672	498346	495	7.15E-138
fucP	104	Intragenic	3286630..3287054	3286819	425	5.55E-36	3286611..3287057	3286781	447	6.53E-187

CE10_0335	-39	Intergenic	370082..370470	370282	389	1.41E-35	370039..370542	370209	504	5.41E-146
yghT	0	Intergenic	3589432..3589897	3589699	466	1.01E-34	3589472..3589821	3589642	350	6.43E-149
yqiG	0	Intragenic	3642370..3642917	3642523	548	2.24E-34	3642410..3642844	3642579	435	6.14E-132
CE10_4295	-781	Intragenic	4363563..4364019	4363746	457	1.26E-33	4364266..4364621	4364436	356	3.76E-104
CE10_0290	0	Intergenic	320797..321310	321125	514	1.69E-32	320918..321276	321088	359	2.39E-86
CE10_1948	0	Intragenic	1971012..1971415	1971220	404	7.04E-32	1970976..1971340	1971146	365	4.95E-210
yebN	-190	Intergenic	2123960..2124427	2124236	468	1.39E-32	2124036..2124415	2124206	380	2.78E-216
yicO	0	Intergenic	4380780..4381184	4380980	405	5.56E-31	4380813..4381260	4380983	448	2.23E-193
lfpB	0	Intergenic	4455502..4456088	4455702	587	1.21E-31	4455601..4455950	4455771	350	1.26E-109
CE10_0275	0	Intergenic	303172..303592	303358	421	1.37E-31	303137..303577	303408	441	6.99E-144
yffB	0	Intergenic	2899415..2899899	2899611	485	1.07E-31	2899399..2899841	2899673	443	4.55E-113
yfgH	0	Intergenic	2938015..2938399	2938226	385	4.07E-30	2938024..2938372	2938194	349	1.66E-286
ompN	0	Intergenic	1605353..1605864	1605666	512	4.66E-30	1605480..1605815	1605645	336	1.00E-122
yobF	-210	Intergenic	2126732..2127148	2126929	417	1.51E-30	2126788..2127160	2126958	373	7.80E-119
yjgN	0	Intergenic	5109874..5110390	5110236	517	2.24E-30	5109939..5110347	5110107	409	6.79E-116
yjeJ	0	Intergenic	4993504..4993937	4993748	434	4.27E-30	4993567..4993930	4993761	364	7.73E-241
yhbX	155	Intragenic	3778601..3779064	3778786	464	9.46E-29	3778615..3779013	3778750	399	9.12E-75
yfgF	0	Intergenic	2937368..2937800	2937603	433	8.10E-29	2937404..2937812	2937643	409	3.31E-126
yihN	0	Intergenic	4624725..4625147	4624948	423	3.25E-29	4624711..4625155	4624991	445	1.68E-84
tsx2	0	Intergenic	4700266..4700782	4700587	517	1.90E-28	4700378..4700758	4700546	381	2.16E-106
CE10_4706	0	Intergenic	4811266..4811714	4811449	449	5.86E-28	4811301..4811756	4811462	456	1.27E-73
dpiB	0	Intergenic	670579..671136	670939	558	1.18E-27	670718..671125	670968	408	1.91E-160
hipB	-478	Intragenic	1739006..1739567	1739206	562	5.88E-27	1739164..1739566	1739321	403	1.77E-130
CE10_1671	0	Intragenic	1702964..1703370	1703173	407	2.02E-25	1702938..1703343	1703105	406	2.64E-121
yihQ	0	Intergenic	4631789..4632250	4632053	462	1.76E-25	4631790..4632294	4632125	505	5.24E-65
gadB	0	Intergenic	1718702..1719278	1719096	577	2.75E-24	1718780..1719174	1718950	395	5.54E-89
yceJ	0	Intragenic	1190063..1190659	1190261	597	1.24E-24	1190104..1190554	1190255	451	4.84E-51

ydjE	741	Intragenic	2071515..2071933	2071734	419	1.47E-24	2071492..2071971	2071823	480	2.56E-108
yjfZ	0	Intragenic	5045764..5046347	5046162	584	3.90E-24	5045991..5046332	5046161	342	8.84E-96
CE10_3786	0	Intragenic	3863969..3864391	3864192	423	6.32E-23	3862491..3862827	3862644	337	1.93E-156
CE10_4709	0	Intergenic	4812986..4813474	4813276	489	2.45E-22	4813018..4813406	4813265	389	6.79E-83
CE10_3570	0	Intergenic	3639848..3640320	3640144	473	9.64E-22	3639939..3640229	3640086	291	2.45E-87
ypdI	0	Intergenic	2808050..2808461	2808249	412	2.78E-22	2808140..2808489	2808306	350	1.33E-261
yegR	0	Intergenic	2432401..2432923	2432753	523	1.99E-21	2432457..2432888	2432721	432	7.56E-115
ompT	0	Intragenic	607490..607945	607749	456	2.04E-21	607592..608028	607728	437	6.42E-48
araE	0	Intergenic	3334405..3334850	3334605	446	2.35E-21	3334471..3334797	3334626	327	1.54E-40
CE10_0264	0	Intergenic	289003..289480	289200	478	2.88E-21	289016..289498	289185	483	1.02E-149
yeiL	313	Intragenic	2560996..2561409	2561216	414	2.69E-20	2561140..2561398	2561222	259	1.36E-51
allB	0	Intragenic	542144..542568	542344	425	4.19E-20	542248..542581	542412	334	2.68E-44
yjdA	0	Intergenic	4926009..4926492	4926209	484	3.01E-19	4926068..4926402	4926254	335	4.39E-125
CE10_4222	0	Intergenic	4299036..4299530	4299227	495	5.03E-14	4298913..4299400	4299083	488	5.06E-53

Table S-7 Overlap of the ChIP binding hits between CFT073 and CE10 (+ D-ser).

CE10 DsdC1 ChIP peak hits	Overlap of ChIP peak hits	CFT073 DsdC ChIP peak hits
CE10_0023	*adiY*	*arcA*
agaC	*agaB*	*aslA*
CE10_0275	*ais*	*asst*
CE10_0290	*araE*	*bglG*
CE10_0366	*bdm/CE10_1671*	*c0322*
CE10_0523	*c0077/CE10_0066*	*c0325*
CE10_0560	*c0080/CE10_0067*	*c0327*
CE10_0724	*c0346/yahA*	*c0363*
CE10_0734	*c0410/CE10_0258*	*c0430*
CE10_0818	*fdeC/CE10_0264*	*c1166*
CE10_0889	*c1618/ymgG*	*c1208*
CE10_1660	*c2348/nmpC*	*c1271*
CE10_1661	*c2383/ysdS*	*c1892*
CE10_1948	*c2815/yfbL*	*c2406*
CE10_2689	*c2897/CE10_2740*	*c3302*
CE10_3076	*c3031/CE10_2939*	*c3692*
CE10_3277	*c3686/kpsF*	*c3694*
CE10_3282	*c3712/CE10_3508*	*c4214*
CE10_3301	*c3766/CE10_3557*	*c4303*
CE10_3398	*c4015/CE10_3786*	*c4498*
CE10_3626	*c4020/CE10_3790*	*c4502*
CE10_3627	*c4423/CE10_4160*	*c4539*
CE10_4229	*c4488/CE10_4222*	*c4920*
CE10_4234	*c4492/CE10_4228*	*clbR*
CE10_4270	*c4739/CE10_4466*	*chiA*
CE10_4295	*c4896/CE10_4613*	*cydA*
CE10_4297	*c4983/CE10_4706*	*fimA*
CE10_4312	*c4986/CE10_4709*	*hcp*
CE10_4363	*c4994/yjbE*	*hlyC*
CE10_4519	*c5088 /ytcA*	*iraD*
CE10_4553	*c5209/ibrA*	*ompC*
CE10_4567	*c5298/CE10_4943*	*papA*
CE10_4598	*c5356/yjgN*	*rpiB*
CE10_4717	*c5426/CE10_5091*	*tdcB*
CE10_4768	*c5435/CE10_5099*	*tssB*
CE10_4870	*caiT*	*yciF*
CE10_4997	*cmtB*	*ydbA_1*
CE10_5101	*cyoA*	*yheE*
CE10_5175	*dpiB*	*yjjQ*
csiD	*dsdC*	
dsdC2	*envR*	

dusA	*fdrA*	
ecpD	*fimE*	
eivC	*frc*	
eivF	*fucP*	
eivG	*gadA*	
epaO	*garP*	
epaS	*hemB/CE10_0335*	
eprH	*iraP*	
espX1	*leuO*	
espX4	*livK/CE10_3982*	
espY4	*nanC*	
eutS	*ompN*	
feoA	*pagP*	
fepE	*pitB*	
fimB	*rfaJ/waaT*	
fimZ	*rfaY*	
focA	*sat*	
fucA	*stpA*	
gadB	*tauA*	
gadX	*tnaA*	
glgS	*waaL/rfaL*	
hlyE	*waaV*	
hutU	*wzx*	
kpsM	*wzy*	
lpfA	*yadN*	
matA	*ycdT*	
mocA	*yebB*	
msbB	*yebN*	
nanR	*yeeN*	
neuA	*yegR*	
neuB	*yfaL*	
neuO	*yfcV*	
prgI	*yffB*	
rcsA	*yghT*	
rfbC	*ygiL*	
setC	*yhbX*	
sfmA	*yhfL*	
sfmC	*yhhZ*	
sfmH	*yhiM*	
sipB	*yiaY*	
sipD	*yicO*	
slp	*yiiE*	
speF	*yjcS*	
tdcR	*yjdA*	
torY	*yjeJ*	
tsx2	*ykgI*	
ttdR	*c0407/rpmE2*	

ubiC	*yliE*	
ugd	*ypdI*	
wbbC		
yahL		
yajR		
ybbW		
ybcK		
ybcM		
ybdN		
ybdO		
ybeF		
ycbQ		
ycbR		
yceJ		
yciD		
ydeK		
ydeP		
ydjE		
ydjO		
yegH		
yehD		
yeiT		
yfdF		
yfgF		
yfgH		
ygcG		
ygeG		
yghJ		
yhaC		
yhaI		
yiaO		
yiaT		
yiaW		
yiiG		
yjbM		
yjiC		
yobF		
yqeI		
yqiK		

Table S-8 Overlap of genes involved in binding sites and genes that were differentially expressed

CFT073 no D-ser	CFT073 + D-ser	CE10 no D-ser	CE10 + D-ser
dsdC	*dsdC*	*dsdC2*	*neuA*
		ypdI	*neuB*
		gadB	*neuO*
		neuO	*rfaY*
			wbbC
			wzx
			waaT
			waaV
			dsdC2